



QY 124 NCLINL---MKTGDDYATSETN-----FIRKIDPOTLETLD 156  
 DB 148 LLMVNTGQIPIKILINLITNGNIALALVYHSGKLLALQEAUKPVVIVKLEDDGLQITG 207  
 QY 157 KVDYSKYVAVNLATSHPHYS-AGNIIINMOTISVDKGRKYVLFKIPSSVPEKKEKSCF 215  
 DB 208 LIDYUKRLTHSF-TAHPKVDVPVTGEMETFGYS-----HTPPYLTYRVIS-----KDG1 254  
 QY 216 KHELVWCIPSKSLLOQSYHSCFITENYIVFTEQPKL---DIVKLATAYINGVWASG 272  
 DB 255 MHDVPPTIS-----EPIMHDFAITETVAIFMLDMPHHPKEMVK-----EKKM1 300  
 QY 273 LSFHKEDKTWFHVDRAKTKKEVSTK-FYTDALVLYHHINAYREDGHVVDIVAYRNSLY 331  
 DB 301 YSFDPITKARCGVLPYAKUDELMIWFLPNCFLFINANAWEEDEVVL-ITCRLENPD 359  
 QY 332 DMFYLLKLDKDFEVNNKLTISPTCKREVPLQYDKDAEVSGLVKPTSATAVKEKDGSI 391  
 DB 360 DMVSKVKEKLENGNELYEM-----RF-----NMKTGSA 389  
 QY 392 YCQPFILCEGTELPFRVADYNGKKYKYVYATEVQWSPVPTKIAKINVOT-----440  
 DB 390 -SQKLSASAVDFPRINFCYIGKKQRYVYGTILUSTAKVTGIIKFDLHAEATGKPMLEV 448  
 QY 441 ----KEVLHMGEDHCWSEPIFVSPDAREDEGVLTICVWSEPNKAPFLLLIDAKTEK 496  
 DB 449 GGNIKGIYDUGEGR-YGSEALIVPRETA-BEIDGCLIFVVDHENTGKS-CVTIVIDAKINS 505  
 QY 497 ELGRATVNV--FMHLDLHGMFIPONDCAET 525  
 DB 506 AEPVAVVELPHRVPGYGFHALFVTEQLEQT 536

## RESULT 7

T49193

neoxanthin cleavage enzyme ncl - Arabidopsis thaliana

N:Alternate names: protein MAA21.150

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Jun-2003 #sequence\_revision 02-Jun-2003 #text\_change 02-Jun-2003

C:Accession: T49193

R:Requester: M. Mueller-Auer, S. Zipp, M. Schaefer, M. Mewes, H.W. Rudd, S. Lemcke, K.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z55018

A:Accession: T49193

A&gt;Status: preliminary

A:Molecule type: DNA

A:Posidnos: 1-538 &lt;PI&gt;

A:Cross-references: EMBL:AL163818; GSPDR:GN06061; ATSP:MAA21 150

A:Experimental source: cultivar Columbia; BAC clone MAA21

C:Genetics:

A:Gene: ATSP:MAA21.150

A:Map position: 3

A:Features: 72/3 56/2, 181/2, 389/3, 372/1, 193/1, 290/1, 291/1, 308/1, 377/2, 417/2, 43

## Query Match

Best Local Similarity 11.0%, Score 310, DB 2, Length 538.

Matches 145, Conservative 7%, Mismatches 213, Indels 144, Gaps 25.

QY 9 KEEHEPEPK-AFVQGGPTWLOGVLLANGMGHITIGTRYNNHWFDDGLALHSPTKN 62

DB 56 RDETP-PVKDLPVHGLPECLNREFVAVGP-----NPKFDAVAGVHWFDDGMIHGVPI 108

QY 63 KNEVYHPSKYIPSYTNYNFIANPIVVSFEETMAYPDPCKNIFAKAFSYLSHTTFFTD 122

DB 109 KDGKATVYSRY-----VKTSRKQEEF-----PFAAKPMKTDGKGRFG 147

QY 123 NCLINL MKTGDDYATSETN-----FIRKIDPOTLETLD 156

DB 148 LLMVNTGQIPIKILINLITNGNIALALVYHSGKLLALQEAUKPVVIVKLEDDGLQITG 207

QY 157 KVDYSKYVAVNLATSHPHYS-AGNIIINMOTISVDKGRKYVLFKIPSSVPEKKEKSCF 215

DB 208 LIDYUKRLTHSF-TAHPKVDVPVTGEMETFGYS-----HTPPYLTYRVIS-----KDG1 254

QY 216 KHELVWCIPSKSLLOQSYHSCFITENYIVFTEQPKL---DIVKLATAYINGVWASG 272  
 DB 255 MHDVPPTIS-----EPIMHDFAITETVAIFMLDMPHHPKEMVK-----EKKM1 300  
 QY 273 LSFHKEDKTWFHVDRAKTKKEVSTK-FYTDALVLYHHINAYREDGHVVDIVAYRNSLY 331  
 DB 301 YSFDPITKARCGVLPYAKUDELMIWFLPNCFLFINANAWEEDEVVL-ITCRLENPD 359  
 QY 332 DMFYLLKLDKDFEVNNKLTISPTCKREVPLQYDKDAEVSGLVKPTSATAVKEKDGSI 391  
 DB 360 DMVSKVKEKLENGNELYEM-----RF-----NMKTGSA 389  
 QY 392 YCQPFILCEGTELPFRVADYNGKKYKYVYATEVQWSPVPTKIAKINVOT-----440  
 DB 390 -SQKLSASAVDFPRINFCYIGKKQRYVYGTILUSTAKVTGIIKFDLHAEATGKPMLEV 448  
 QY 441 ----KEVLHMGEDHCWSEPIFVSPDAREDEGVLTICVWSEPNKAPFLLLIDAKTEK 496  
 DB 449 GGNIKGIYDUGEGR-YGSEALIVPRETA-BEIDGCLIFVVDHENTGKS-CVTIVIDAKINS 505  
 QY 497 ELGRATVNV--FMHLDLHGMFIPONDCAET 525  
 DB 506 AEPVAVVELPHRVPGYGFHALFVTEQLEQT 536

## RESULT 8

S76206

hypothetical protein - Synchocystis sp. (strain PCC 6803)

C:Species: Synchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S76206

R:Requester: T. Sato, S. Kotani, H. Tanaka, A. Asamizu, F. Nakamura, Y. Miyajima,

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res 3, 109-116, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis

s.

A:Reference number: S74322; M0109702101; PM019905241

A:Accession: S76206

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-480 &lt;KAN&gt;

A:Cross-references: EMBL:D90914; GB:AH001349; NID:q1654477; PID:HAAL18465.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

## Query Match

Best Local Similarity 23.4%, Score 280, DB 2, Length 480.

Matches 113, Conservative 90, Mismatches 210, Indels 70, Gaps 19.

QY 5 FNRKKEPEPKAEVQGGPTWLOGVLLANGMGHITIGTRYNNHWFDDGLALHSPTKN 64

DB 14 FSNNAVEELPTETLFLTSQGLFALIKSLIYRPGKRSQSLKGVHWFDDGLALHAFGE 74

QY 65 GEVYHPSKYIPSYTNYNFIANPIVVSFEETMAYPDPCKNIFAKAFSYLSHTTFFTD 123

DB 74 GDAELFYVLCGVALLLEFAGSYFCHQDMV--GVVAVVWFLTGVVAVFPAH 129

QY 124 CLINMIGLGYVAISHENFIRKIDPOTLETLDKRVY-SKYVAVNLATSHPHYS-AGN 181

DB 130 --TSVMALPDLTWLWAGHPYALDLENLATIGTFPGGAFQGPQSFSAHPIDAVTDEI 187

QY 182 LNMGTISVKGPTKYVLFPSVSVFEKKEKSTETIIFVVSFESPSLTPSYVFFTH 241

DB 188 FSIQ-----VDFQNLNLYPLDQSGNLKHPFL-----KLSPTPTPNSPMA 229

QY 242 ENYIVFTEPEKEDLVKLATAYINGVWASGSEFHEEKILWDFVLRKAKASIKETID 301

DB 230 GPYLVLEPLDPTITNQESL-----LGNVAYALQAWAGQTEKAVVFPQSLTETVSGTTE 285

QY 402 ALVLYHHINAYE-FHCHVWFVDIVAYRNSLYNGEYUKKLEGEFVANNFESLSPFEPVV 460

DB 286 SWFGWYHNGCELDDGNGVRLGDFVRENFA-----QINYEYLRVET-----G 426



A:Reference number: 221500

A:Accession: T34293

A:Status: preliminary; translated from gb/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1760 <ML>

A:Cross-references: EMBL:U53341; PIDN:AAC6q106.1; GSPDR:GN00028; CFSP:F49E10.2

A:Experimental source: strain Bristol N2; clone F49E10

C:Genetics:

A:Gene: GSP:F49E10.2

A:Map position: X

A:Introns: 51/1, 42/3, 176/3, 235/3, 332/2, 514/1, 543/2, 505/3, 677/1, 732/3

#### Query Match

Best Local Similarity 9.5%; Score 269; DB 2; Length 790;

Matches 136; Conservative 91; Mismatches 221; Indels 154; Gaps 28;

QY 5 FNRKKEHPEPIKAEVQGL-----PTWLGVL-----RNGPMHTIGDTKYNHWD 52

DB 20 FNTHLGHPAAMDGKYRELYCPSKNIPKWLGGYHLCQSASYGNSSAPEGE-KLNHMD 78

QY 53 GLALLHSFTKNGEVYYSKYLRS-----DTYCNLI-----EANRIVVSEFQT 95

DB 79 AIGAVGSHVNSGVVSQAQYFAPRYK IWEFTDRNMSKASVPWAGSDYNLTAMSRWEQ 138

QY 96 M-AYDPCKNIFAKAFSYLSHTIPEFDNCLINIMKTGGDYATSETNFI--KKIDPQTL 152

DB 139 VPANPDSAR-----FHPN--LDFWKVGNRIVAGTEAPYVWGVYEDVRLT 180

QY 153 ETLKVDYDK-----YVAVNLATSHPHYDSAGNI-----LN-----MGTSI 188

DB 181 QKFLPFKPEKNDIFSTPRHTMIPISMAL-HERNDAGTIGWFSAMNFEEORFTQGIT 239

QY 189 VDKGRTKVFLFKIPSSVPEKCKKCKHLEVVCSIPSRSLQPSVYHSEFGENIV-- 246

DB 240 VDTNCGVRVGLIDYGVWQ---INAGSSNEY---IGDKLL-PGYIHSITSTENFILP 292

QY 247 -----FIEOPFKLDIVKLATAYIRGVNNSCLSPHKEDKTHFVDRKTKKEVST 296

DB 293 ITSLLINCKFKPEPLN---NVRSAIQKGLMG---MDEYDMVPMRFLFNKTKLEFTTS 346

QY 297 K--EYTDALVLYHINAYE EDGHVVFEDIWAYEDNSLYDMFYLKLDKDKDFEYNNKLTSP 353

DB 347 KPLEVFSMFVTHOLNAFEADDGNFVADVMVYDSDHP---YVKYPTDFELTKQLPSTA 402

QY 354 TKCFEVPVPLQYDKDAEVSGLVLPSTASATKEKGSYICQPEILCEGIELPRVNYDNG 413

DB 403 RVLRFET--LDSKKGRVMYNYLVPOETIAA-----DFPQFNHNYEQ 440

QY 414 KKYKVVYVATEVQWSPVPTKIAKLNV-----QTKVYLHWGEDHCWPSFIEVP 460

DB 441 FPGVWGYLVQHPFAS--GNSITFINVPPAGNPNEPFAPTTIVH-----FWPVL 490

QY 441 SPDAFFPHVAVLTGVVVSFPNKAPFILLDAKTEKLGPAFVNVEMHLDHGMFQND 520

DB 491 KPDSKKKEGVGLIWPGLVTAENKT-VLIVTAEKMWELGRAVVFISIPFSFHNFPSKKD 549

QY 521 LG 522

DB 550 LG 551

#### RESULT 12

AG2417

hypothetical protein all4895 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AG2417

R:Kurata, T.; Nakamura, Y.; Wolk, C. P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iizuka,

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG2417

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1472 <OR>

A:Cross-references: Gb:BA000179; PFIN:BAH70594.1; PID:q17134033; GSPDR:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4895

#### Query Match

Best Local Similarity 9.4%; Score 264.5; DB 2; Length 472;

Matches 109; Conservative 72; Mismatches 192; Indels 107; Gaps 17;

QY 20 VQSGPLPWLGSLVFNPGIPGMIHIGDTKYNHWFHETALHSFIPKKNFVYVPSKYLPSDLY 79

DB 32 TSNIPGSLGTLYPNCPAPIFPGGMLAGHWFHETGAILAVNFTDAGVIAITYVQIVQY 91

QY 80 NCNIFANPIVVSFPGTMAYPDPCKNIPAKAFSVLSHTDEPTFNTINIMKTGGDYAT 139

DB 92 QVENTADKELYGVG-MTAPGVVWQWRKPIKNAANTSVLALPDKLLALWEGGRVY 146

QY 140 ETNFIPIPTGTLFTLTKVPSKYVAVNLATSHPHY-----GSA-NININMAFSLVPRKIKY 196

DB 147 -----ALDQTLPTWGEDNLGG--LTNGLTYSARYKQDPQTKETPNFGISLNAK 195

QY 197 VLPKIPSSVPEKCKKCKHLEVVCSIPSRSLQPSVY-----HSEFGENIVFV 248

DB 196 -----LNVYKSDSTGKTIORSAYSLDGLPLIHDFVLGAKYLIFP 244

QY 249 EGPFKLGI--VKLATAYIRGVNNSCLSPHKEDKTHFVDRKTKKEVSTREYTDALVLY 406

DB 235 ISPVRLANLPVLLGTS-----NYSDSMQWRPELGATQILVIDRESLSLVS RGETETWYQW 288

QY 307 HHINAY--EECHGVVFDIVAYEENSLYDMFYLKLDKDFEYNNKLTSTPTCKRFVVPLOYD 465

DB 289 HFANGYVDASGAVLDIARY-----KDFQTN-----QVL 417

QY 466 KDAEVSNI-VKLTPTSATAVK--EFGSTVQGPPEIDFGFIFIPVNVYVW-KKKYKVVY-- 420

DB 318 KEVATGOTHTPATSTLSPVNLHPGSKVAALFGLLPHCFEFDHVPVQNVQASRYTMSG 477

QY 421 ---AIEVQWSKVPKIAKLVJIKVLH-WCHDHQWPSFEDFVPSIDAEEDGCVLLVCV 476

DB 378 FROGTDISQELLNF-TACFDHKTOTFTIAYVGENCYSPSEPIIA-----DNWVLVYV 427

#### RESULT 13

T51936

probable 9 cis epoxycarotenoid dioxygenase [imported] - potato

C:Species: Solanum tuberosum (potato)

C:date: 20-Nov-2000 #sequence\_revision 20-Nov-2000 #text\_change 20-Nov-2000

C:Accession: T51936

R:Burridge, A.; Taylor, J. B.; Thompson, A.

submitted to the EMBL Data Library, March 2000

A:Description: potato putative 9 cis epoxycarotenoid dioxygenase 1 cDNA.

A:Reference number: 225874

A:Accession: T51936

A:Status: preliminary; translated from gb/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-604 <OR>

A:Cross references: EMBL:AJ26244; FIDN:CAH76920.1

C:Genetics:

A:Gene: need1

#### Query Match

Best Local Similarity 9.3%; Score 264; DB 2; Length 604;

Matches 128; Conservative 94; Mismatches 219; Indels 154; Gaps 27;

QY 1 MEIIFPNKKEEPEPIKAE-----VQGLPTWLGKLVLLKNGNG 48

DB 104 VEGALTKHELEFIPKPTADFVFGVLSGNFAVFPNFW-QSLIVTKRTKGVGVVYVNRKAN 164

QY 39 MHTIGDTKYNHWFHETGALLHSFTEKNGEVYYSKYLRSIYNCTNIPANKIVVSEFETMAY 98



Db 164 PLPEPTAGPHFFGCGVHAGCKKASASYAFETETEPF-----VQEKALGRPVF 214  
 QY 99 PDPKNI-----FAKA-FSYLSHT-----IPETTNL-----I 176  
 Db 215 PKATCELGCHGSLAPLEFAPOLGLDHSRGTQVANAGVYVFNRLAMSPDOLFYHV 274  
 QY 127 NIMKTGMYVATSETNFIKTPRTFTIHKVNSKYVAVNATISHQKYS-AGNLLNWG 185  
 Db 276 KVITFG-----LAKGFFLQGLKSLN-IAHPKLDPVGELFALS 315  
 QY 186 TSIVKSPKTKYVLEKIPSSVPEKKKSKFKHLEWVSIIPSPSLDPSVYHSGTIPNY 245  
 Db 316 YDVIOKPKLVKVF-----SKNGEKS--NDVEI-----PVEDPTMMHDAITEKEV 359  
 QY 246 VFIEOPKLDIVKATAIKGVNWSCLSPHKEDTKWFHFVDKTKKYSTKPY-TDALV 304  
 Db 360 LIPO-----QVFMKSEIRG---GSPVYVYKKNKVSFGILDKYKAGSNLKWVDPDF 412  
 QY 305 LYHHINAYEE-DGHVVEDIVAYKNSLYMFLKLLKLEFVNKKLISIPCKKGFVVP 361  
 Db 413 CFHLNNAWEETDEIVVIGSCMTPPDSEFN-----ETDELSFVL-----453  
 QY 362 LQYKEDARVNSNKLKLSATAVEK-PSIVYQPELLTQELFPPVNVYNNKKYKYV 429  
 Db 454-----SEIRLNLTGSTRKAKIENDEQV-----NLEAGVNVNKKIUSPKTQYAY 498  
 QY 421 ATEVGSVPVTKIAKLVQKFEV--LHWGECHWPSEPIFVP-SPAREGEFGVLTQVV 477  
 Db 499 LAIAPPWPKVSPAKVILFTTEVKKPIYV-LNKYGRPLRTPPKVNSKRFHDSYIL-APV 556  
 QY 478 VSEKNKAPFLILDAKTEKELGRATVNVPMH--LQHGMEFIPNQL 521  
 Db 557 HDEKFWTSQIVNMTIKL--EATVKLPSPVYGEHGFINANDL 600

## RESULT 14

E96812  
 C:Species: Arabidopsis thaliana  
 C:Date: 02-Mar-2001 #sequence\_revision 02 Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96812  
 R:Theologis, A.; Eckert, J.R.; Paim, C.J.; Fiedor-Spiel, N.A.; Kaul, S.; White, O.; Altschuld, C.W.; Chung, M.K.; Gnan, L.; Conway, A.R.; Conway, A.F.; Greasy, L.H.; Lewis, R.; ausen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Jandas-De-Bevoise, C.; Khan, S.; Khaykin, E.; Kim, S.; Rizzo, M.; Rooney, T.; Rodley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, P.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: AB6141; MUID:21016719; PMID:11130712  
 A:Accession: E96812  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-667-Stop.  
 A:Cross-references: GB:AE005174; NID:98065443; PIRN:AAF17971; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F3F9.10  
 A:Map position: 1

Query Match 9.38, Score 261.5, DB 2, Length 657;

Best Local Similarity 22.94, Pred. No. 1.1e-10;

Matches 130, Conservative 87, Mismatches 229, Indels 121, Gaps 26;

QY 4 IFNENKEHPPIKAE-----VGGQLPTWGLGVLLKNGCG-MH 40  
 Db 163 MISERRRHPKPTADPAVGIAGNEFFVPKPVVHNLPLVGIPELQGVYVKNAGNPLH 222  
 QY 41 TIGDTKYNHMFQIALLSHTTPKNGFVYKPSYKRSITYNKCNIEANPIVVS-EPFGTMAYP 99  
 Db 223 F-PVSCHHLEFGCGVHVAFFETNGSVSW-54-----IFINAGVQ-EEGSHVVF 271

QY 199 DPKKNI-----EAKAPSYLSHTIPEPID-NCLINIMTKGVVY-----ATSETNF---IR 145  
 Db 272 KATCELGCHGSLAKIMLPNLPDPAVHPTGGGAVANAGLVYFPHLLAMSTGTYHYV 331  
 QY 146 FVQGTETFTKIVSYFVAVNLAISHRYD-SAGNLLNMKTSIVKSGTKVLEKTPSS 204  
 Db 332 VIGTGGTETSPPTETIQQKSTM-IAHPFTETPETHFALSVIVSKFYKYEPF-----885  
 QY 205 VPEKPFPSKPKFEVY--LQPSLSLQPSYHFFIENIVIFIPFPELDAVELAIAVI 264  
 Db 486 ISDEKSD-----VEIP-----LLQPTMHHFALPFPNVVPIPQ-----QVVERPEML 440  
 QY 265 PGVNWASCLSPHKEDTKWFHFVDKTKKYSTKPY-TDALVLYHHINAYEE--DGHVVP 420  
 Db 431 PG---GSPVYDEKKKSFGLNKNKADASSIONIEVPDQCFCHLNSWEEETDEVV 487  
 QY 321 DIVAYPNSLYDMFLYKLLDKDFEVNNFLTSIPCKPFVVPICQYKDAEVSGLVPIPS 480  
 Db 488 GSCMPPESIFN-----EHDETQSVL-----SEIRLN-----516  
 QY 391 ALAVKKEGDSYQCPPICTF-GIFLPPVNYVYNNKKYKYVYVATFVQVSPVTKIAYLNQ 479  
 Db 517-----FIDESPPFPVVISFVNIFAKVNPNLQPTPTFYAYIAITFWPKVSGFAKVL 570  
 QY 440 IKEVLH--WGEDCHWPSEPIFVPSAPAEDEGVLTQVVSVPNKPAPFLILDAKTEK 497  
 Db 571 IGRPPYVYCGK-YGDEPLFISQCG-EEUGGVYVNVHIDREFVASE-LQLINAVNMKL 627  
 QY 498 LGPATVNVEMHLDLHGMFIPQNDIGAE 524  
 Db 628 EATVILSERVYQFHEIFSEFIDFQ 654

## RESULT 15

JC4324  
 C:Species: pseudomonas paucimobilis  
 C:Date: 08-Jan-1996 #sequence\_revision 08-Feb-1996 #text\_change 17-Mar-1999  
 C:Accession: JC4324  
 P:Kamada, S.; Suburi, Y.  
 Biosci. Biotechnol. Biochem. 59, 1866-1868, 1995  
 A:Title: Purification of a lignocellulose-degrading, beta-dioxxygenase isozyme from a pseudomonas strain J-4324, MUID: 4606831, PMID: 8534977  
 A:Reference number: JC4324  
 A:Molecule type: DNA  
 A:Accession: JC4324  
 A:Residues: 1-489-KAM>  
 A:Experimental source: TMV1009  
 A:Note: The translation initiation codon AUG is not given in this paper  
 C:Comment: This enzyme is significant in the metabolism of dimeric lignin compounds.  
 C:Genetics:  
 A:Gene: lsdb  
 C:Keywords: oxidoreductase

Query Match 8.79, Score 245.5, DB 2, Length 489;

Best Local Similarity 22.48, Pred. No. 9.4e-10;

Matches 121, Conservative 76, Mismatches 211, Indels 131, Gaps 22;

QY 19 EVQGLPIWLAGVLLPNPQPMHIDGFKYHWEGLALHSHFPPKRVVYPSKYLRSDT 78  
 Db 26 EVGSHTPAQITGIFPKVYHVAQDEHPHPEFDHPPHNDQMSLVSEFHHCKILDERQKYACILK 85  
 QY 79 YNCHNEANSIVVSEFGTMAYTIVYENIFAKATSVLSHTIETDRLI 81 128  
 Db 86 WKVPKACK--SLF3-AYENP-----TMTASVQGMTPGIANHW 122  
 QY 129 MKIGUDGYVATSENF-PEKIDPQILLELDKVDYSKAVANLAISHRYDS-AGNLLNMKTS 187  
 Db 123 VVHAKKLYAMKEUSFQIMELLETESYTNFQCKLKNQTSNAHIDIVTGNCFQYA 182  
 QY 188 IVDKPTKYVLFKIPSSVPEKKKSKFKHLEWVCSIPSSL-----GPSYK---HSPG 249  
 Db 193 AGCLLIGCSYFELD-----FACMLLELELFQVYVYGMHHPYD 220





```

DR pfam: PF00128; alpha-amylase; 1.
DR pfam: PF00395; SLH; 3.
DR pfam: PF02806; alpha-amylase-C; 1.
DR pfam: PF02903; alpha-amylase-N; 1.
DR pfam: PF02922; isomylase-N; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01072; SLH_DOMAIN; 3.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
KW Multifunctional enzyme; Glycoprotein.
FT SIGNAL 1 35
FT CHAIN 36 1861 AMYLOPULLULANASE.
FT DOMAIN 428 1018 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1157 1248 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 628 628 BY SIMILARITY.
FT ACT_SITE 657 657 BY SIMILARITY.
FT ACT_SITE 734 734 BY SIMILARITY.
FT DOMAIN 1681 1739 SLH 1.
FT DOMAIN 1740 1803 SLH 2.
FT DOMAIN 1804 1861 SLH 3.
FT CONFLICT 1734 1734 D -> E (IN REF. 1; AAB00841).
SQ SEQUENCE 1861 AA; 206104 MW; 8623070EA548574 CRC64.

Query Match
Best local Similarity 4.38; Score 121; Nr 1; Length 1861;
Matches 117; Conservative 91; Mismatches 224; Indels 146; Gaps 30.

QY 23 QLPTWLGVLRLNGPG-----MHTIGDKYNNHW-----FDGLAL-LHSFTFKNGEV 67
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
931 GAPS:---NVVYTSNGKVDLSWLSGATGYNLYKSSVEGLYKIASNVTTTFEDANV 987
QY 68 YRSKYLRSDTYNCINIFANRIVSEFG-----TMAYDPCKNIFAKAFSYLSHTIPEFTDN 123
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
988 TNGLYKYVA-----ISAIDELNESCISNDAYVP-----AYPIGWGN-LTQVSDN 1033
QY 124 CLINIMKTGDYATSETNFIKIDPOTLETLDKVDYKYVAVNATSPHVDISA-GNTL 182
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1034 HLIGVDKPTDIYAEVWADGLTNGTGGPNMIAQLGY-KYVSTV-----YDSVVGSVY 1086
QY 183 NMGTSIVDKGRTKYVLFKIPSSVPEKEKKSCFKHLEVCSIPSRLLQPSYYHSGF--- 239
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1087 NSVYGVDDSGFT-WNNAQYVGDIGNDQYKASF-----TPDK-IGQWEVLMRFSNQ 1136
QY 240 ----ITENYIVFIEQPKLDIVKLATAYIRGVNWSCLSFHKEDKTWFHFVDKTKKEVS 295
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1137 GDMWTTSTLSFYVWPSD-DLIKPTAPYLNQPTES-----SRVSTFW----- 1178
QY 296 TKFYTDALVLYHHINAYEEDGHVVDIVAVRDNLSYD-----MFYLLKLLDKDFEV 345
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1179 -NPSTDNVGIYDY-EIYPSDGG-TFNKIATVSNEVINYIDTSVINGVTYNYKVVAVDLSF 1235
QY 346 NKKLTSIPTCKRFVVPVLPQ-----YKDAEVSGLNV-KLPT-----SATAVKEKDG 390
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1236 NKTESVNTIKPDVVPVPIKFIENVTVDPDTPCA---VNLAGTFPNATWDP-SAGQMTKIDNN 1292
QY 391 IYCQPEILCEGIELPRVNYDNGKKYKYVYATEVQSPVP-----TKIAKLNVQ 439
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1293 TYSITLIDETQI-----EKYARGS---WDKVEKDEYGNFASNRKVTIVNOG 1339
QY 440 TKRV:HWFFRHWNPSEPTFVPSDAPAEEREGVYLTCCVVSSEPNKAPFLILDKATFKE-- 497
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1340 NNEMTINQTVYWRDIPFIYSPSSNMTVDSNISITMEVKGNTYKGAKVTI-NGDSFVQDK 1398
QY 498 -----LGRATVNVEMHLDLHGFMFIPNDLGAETIE 526
DB 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1399 NGVPTKVSLSNYGVNKKIKIHVEPNDSGVYNGNDGRIE 1436

RESULT 2
PEX_HUMAN
ID PEX_HUMAN STANDARD: PRT: 749 AA.
AC P76562; Q00678; Q98827; Q13646; Q91032;
DI 01-NOV-1997 (rel. 35; Created)
DT 01-NOV-1997 (rel. 35; Last sequence update)

```

```

DI 15-JUN-2002 (rel. 41; Last annotation update)
DE Phosphate regulating neutral endopeptidase (EC 3.4.24.-)
DE (Metalloendopeptidase homolog PEX) (X-linked hypophosphatemia protein)
DE (HYP) (Vitamin D-resistant hypophosphatemic rickets protein).
GN PHEX OR PEX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS HYP ARG-85; LEU-514; ARG-579 AND
RP PRO-651.
RX MEDLINE=97343325; PubMed=9199930;
RA Francis F., Strom T.M., Hennig S., Boeddrich A., Lorenz R.,
RA Brandau O., Mohrke K.L., Cagnoli M., Steffens C., Klaus S.,
RA Borzym K., Pohl T., Oudet C.L., Econs M.J., Powe P.S.N., Reinhardt R.,
RA Meitinger T., Lebrach H.;
RA "Genomic organization of the human PEX gene mutated in X-linked
RT dominant hypophosphatemic rickets.";
RL Genome Res. 7:573-585(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97232252; PubMed 9077527;
RA Beck L., Soumounou Y., Martel J., Krishnamurthy G., Gauthier C.,
RA Goodyer C.G., Tenenhouse H.S.;
RT "Pex/PEX tissue distribution and evidence for a deletion in the 3'
RL region of the Pex gene in X-linked hypophosphatemic mice.";
J. Clin. Invest. 99:1200-1209(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Bone;
RX MEDLINE=97343443; PubMed=9199999;
RA Guo R., Quarles L.D.;
RT "Cloning and sequencing of human PEX from a bone cDNA library:
RT evidence for its developmental stage-specific regulation in
RT osteoblasts.";
RL J. Bone Miner. Res. 12:1009-1017(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Lipman M.L., Panda D., Henderson J.E., Shen Y., Goltzman D.,
RA Karaplis A.C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=97224400; PubMed=9070861;
RA Griff M., Mumm S., Waeltz P., Mazzarella R., Whyte M.P.,
RA Thakker R.V., Schlessinger D.;
RT "Expression and cloning of the human X linked hypophosphatemia gene
RT cDNA.";
RL Biochem. Biophys. Res. Commun. 231:635-639(1997).
RN [6]
RP SEQUENCE OF 4-641 FROM N.A.
RX MEDLINE=96024647; PubMed=7550339;
RA Francis F., Hennig S., Korn B., Reinhardt R., de Jong P., Poustka A.,
RA Lebrach H., Powe P.S.N., Goulding J.N., Summerfield T., Mountford P.,
RA Read A.P., Popowska E., Pronicka E., Davies K.E., Orlandi J.L.H.,
RA Econs M.J., Nesbitt T., Brezner M.K., Oudet C.L., Pannetier S.,
RA Hanauer A., Strom T.M., Meindl A., Lorenz R., Cagnoli M.,
RA Mohrke K.L., Murken J., Meitinger T.;
RT "A gene (PEX) with homologies to endopeptidases is mutated in
RT patients with X-linked hypophosphatemic rickets. The HYP
RT Consortium.";
RL Nat. Genet. 11:130-136(1995).
RN [7]
RP SEQUENCE OF 1-116 FROM N.A.
RA Waterston R.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [8]
RP VARIANTS HYP TYR-85; CYS-166; SER-252; ILE-254 AND VAL-579.
RX MEDLINE=97260404; PubMed=9106524;
RA Helm T.A., Huang X., Kukel L.M.,
RT "Mutational analysis of the PEX gene in patients with X-linked
RT hypophosphatemic rickets.";

```

Am. J. Hum. Genet. 60:790-797 (1997)  
[9]  
PN VARIANTS HYP SER-77; PRO-138; LEU-534 AND ARG-579.  
RX MEDLINE=97252387; PubMed=937956;  
RA Rowe P.S.N., Oudet C.L., Francis F., Sinding C., Pannetier S.,  
RA Econs M.J., Strom T.M., Mettinger T., Garabedian M., David A.,  
RA Macher M.-A., Questiaux E., Popowska E., Fronicka E., Read A.P.,  
RA Mokrzycki A., Glorieux F.H., Dreznier M.K., Hanauer A., Levrach H.,  
RA Goulding J.N., O'Riordan J.L.H.;  
RT "Distribution of mutations in the PEX gene in families with X-linked  
RT hypophosphataemic rickets (HYP).";  
RL Hum. Mol. Genet. 6:559-549 (1997).  
RN [10]  
RN VARIANT HYP PRO-555  
RX MEDLINE=98439582; PubMed=9768646;  
RA Econs M.J., Friedman N.F., Rowe P.S.N., Speer M.C., Francis F.,  
RA Strom T.M., Oudet C.L., Smith T.A., Nimomiya T., Lee R.E.,  
RA Bergen H.,  
RT "A PEX gene mutation is responsible for adult-onset vitamin  
RT D-resistant hypophosphataemic osteomalacia: evidence that the disorder  
RT is not a distinct entity from X-linked hypophosphataemic rickets";  
RL J. Clin. Endocrinol. Metab. 83:3450-3462 (1998)  
RN [11]  
RN VARIANTS HYP P-317; L-544; K-579; K-621; N-680 DEL; P-740; Y-741 AND  
RP K-749.  
RX MEDLINE=98439610; PubMed=9768674;  
RA Dixon P.H., Christie P.T., Weeding G., Trapp B., Grieff M., Holm L.A.,  
RA Gertner I.M., Schmidtke T., Shah R., Shaw N., Smith C., Liu C.,  
RA Schlessinger D., Whyte M.P., Thakker R.V.;  
RT "Mutational analysis of PEX gene in X-linked hypophosphataemia";  
RL J. Clin. Endocrinol. Metab. 83:3615-3623 (1998)  
RN [12]  
RN VARIANTS HYP S-80; F-142; G-237; G-530; I-573; S-733 AND W-746  
RX MEDLINE=99368844; PubMed=10439971;  
RA Filisetti D., Ostermann G., von Bredow M., Strom T.M., Filler G.,  
RA Ehrlich J., Pannetier S., Garnier J.-M., Rowe P.S.N., Francis F.,  
RA Hollmann A., Harbauer A., Econs M.J., Guder C.L.;  
RT "Non-random distribution of mutations in the PEX gene, and  
RT under-deleted missense mutations at non-conserved residues";  
RL Eur. J. Hum. Genet. 7:615-619 (1999)  
RN [13]  
RN VARIANTS HYP F-85; P-141; V-341 DEL; F-567; K-680 AND Y-693.  
RX MEDLINE=20202840; PubMed=10737991;  
RA Tyysimä H., Kaitila I., Naantol-Salonen K., Ala-Huohala M.,  
RA Alitalo T.;  
RT "Identification of fifteen novel PEX gene mutations in Finnish  
RT patients with hypophosphataemic rickets";  
PL Hum. Mutat. 15:383-384 (2000)  
RN [14]  
RN VARIANTS HYP ARG-160 AND ASN-444 INS.  
RX MEDLINE=20451419; PubMed=11004247;  
PA Sato F., Tajima T., Nakae T., Adachi M., Asakura Y., Tachibana Y.,  
PA Sawa S., Katsumata N., Tanaka T., Hayashi Y., Abe S., Murashima M.,  
PA Okuhara K., Shinohara N., Fujieda K.;  
RT "Three novel PEX gene mutations in Japanese patients with X-linked  
RT hypophosphataemic rickets";  
RL Pediatr. Res. 48:536-540 (2000).  
CC -1- FUNCTION: PROBABLY INVOLVED IN RONE AND DENTIN MINERALIZATION  
CC AND RENAL PHOSPHATE REABSORPTION.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).  
CC -1- TISSUE SPECIFICITY: LYMPHOCYTE AND FETAL BRAIN; NOT IN ADULT  
CC BRAIN, PLACENTA, SKELETAL MUSCLE, AND PANCREAS; NOT IN ADULT AND  
CC FETAL HEART, LUNG, LIVER, AND KIDNEY.  
CC -1- DISEASE: DEFECTS IN PEX ARE A CAUSE OF X-LINKED HYPOPHOSPHATEMIC  
CC RICKETS (HYP). A DOMINANT DISORDER CHARACTERIZED BY IMPAIRED  
CC PHOSPHATE UPTAKE IN THE KIDNEY, WHICH IS LIKELY TO BE CAUSED BY  
CC ABNORMAL REGULATION OF SODIUM PHOSPHATE COTRANSPORT IN THE  
CC PROXIMAL TUBULES. CLINICAL MANIFESTATIONS INCLUDE SKELETAL  
CC DEFORMITIES, GROWTH FAILURE, CRANIOSYNOSTOSIS, PARAVERTEBRAL  
CC CALCIFICATIONS, PSEUDOFRACTURES IN LOWER EXTREMITIES, AND MUSCULAR  
CC HYPTONIA WITH ONSET IN EARLY CHILDHOOD. X-LINKED HYPOPHOSPHATEMIC  
CC RICKETS IS THE MOST COMMON FORM OF HYPOPHOSPHATEMIA WITH AN  
CC INCIDENCE OF 1 IN 20000.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13.  
CC -1- DATABASE: NAME: EMBL; WWW: <http://data.mbi.mcgill.ca/pep.html>.  
CC  
CC This SWISS Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcements/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y10196; CAA11258.1; -  
DR EMBL: U75645; AAB47749.1; -  
DE EMBL: U82970; AA24487.1; -  
DE EMBL: U87384; AAI47562.1; -  
DE EMBL: Y08111; CAA69326.1; -  
DE EMBL: Y08112; CAA69326.1; JOINED.  
DR EMBL: Y08113; CAA69326.1; JOINED.  
DR EMBL: Y08114; CAA69326.1; JOINED.  
DE EMBL: Y08115; CAA69326.1; JOINED.  
DE EMBL: Y08116; CAA69326.1; JOINED.  
DE EMBL: Y08117; CAA69326.1; JOINED.  
DR EMBL: Y08118; CAA69326.1; JOINED.  
DE EMBL: Y08119; CAA69326.1; JOINED.  
DE EMBL: Y08120; CAA69326.1; JOINED.  
DE EMBL: Y08121; CAA69326.1; JOINED.  
DE EMBL: Y08122; CAA69326.1; JOINED.  
DE EMBL: Y08123; CAA69326.1; JOINED.  
DE EMBL: Y08124; CAA69326.1; JOINED.  
DE EMBL: Y08125; CAA69326.1; JOINED.  
DE EMBL: Y08126; CAA69326.1; JOINED.  
DE EMBL: Y08127; CAA69326.1; JOINED.  
DE EMBL: Y08128; CAA69326.1; JOINED.  
DE EMBL: Y08129; CAA69326.1; JOINED.  
DE EMBL: Y08130; CAA69326.1; JOINED.  
DE EMBL: Y08131; CAA69326.1; JOINED.  
DE EMBL: Y08132; CAA69326.1; JOINED.  
DE EMBL: Y08133; AAB42219.1; -  
DE EMBL: U81167; AAB42219.1; JOINED.  
DE EMBL: U81168; AAB42219.1; JOINED.  
DE EMBL: U81169; AAB42219.1; JOINED.  
DE EMBL: U81170; AAB42219.1; JOINED.  
DE EMBL: U81171; AAB42219.1; JOINED.  
DE EMBL: U81172; AAB42219.1; JOINED.  
DE EMBL: U81173; AAB42219.1; JOINED.  
DE EMBL: U81174; AAB42219.1; JOINED.  
DE EMBL: U81175; AAB42219.1; JOINED.  
DE EMBL: U81176; AAB42219.1; JOINED.  
DE EMBL: U81177; AAB42219.1; JOINED.  
DE EMBL: U81178; AAB42219.1; JOINED.  
DE EMBL: U81179; AAB42219.1; JOINED.  
Query Match 4094, Score 112.5, DB 1, Length 749;  
Best Local Similarity 20.18, Pred. No. 1.6;  
Matches 97, Conservative 70, Mismatches 176, Indels 139, Gaps 23;  
QY 33 LENGTH=EMBL:U81176;GALHSHFIPANGLVYKSLKSLDYNKLNKRNKRVVSE 92  
DE 236 LKNETEAYSYFALAKYKFWETAVLL...GANSSEAEHCKPVSVEIKPILALM 286  
QY 53 EGMATYDQAKNITAKA...EYLSHSHIIFEFHNLINIMKIGIAYVASEIDHFK 146  
DE 287 -----PHENPTSEAMYNKMNISELSAMIQP-----DW...LQYIKK 320  
QY 147 -IDGCTLETTIKVYSKYVAVNLAISHPHYSACNITINMTSTVDKGTYYVIER-TPSS 204  
DE 321 VDPDPLYPHIKDISPSNVVVVPV---PQVFK--DIFRIGSFKKTKIANYLVMPVYSR 374  
QY 205 VDEKKKSKCFKHEVVCSTPSRELLOPSYHSPITEN---YIV---FTEQPKLIDVK 258  
DE 375 (PNI)SPSPQ--VFWLPPSPVIVGSTITLIPQWLKQVNFIESALPYVVGKMFVLYVGEKKE 433  
QY 259 LATATPQVNNASCLSFHKEKTKWFHFVQCKKEVSTKE-----YTDALVLVHHIN 310

```

DB 434 MMEELIVEGVWAFIDMLEKENE-----WMDACTKPKAKKAPAVLAKVGYPPIIMNDTHVN 489
QY 431 AYEDGHVWVDIVAYRONSILY-----DMFYLLKDKDFEVNNKLTISPTCKRFVPL 362
DB 430 --EDLKAIFSEADYFGNVLTGKYLAQSDFFWLR-----AVPKTEWFTNPT 535
QY 463 QYDKDAEVGSLVKLPTSATAVKKGSIYQPEILCEGIEIPRV----- 407
DB 536 TVNAFYASINQIRFPAG-----ELUKPFW-----GTEPRSLSYGAIGVIVGHEPT 583
QY 408 --NYDNGKKY-----KYVYATEV-----QWSPVPTKIATLNVOTKEVLHWG 447
DB 584 HGFUNGKKYDKNGLDPMWTESEKFKETKCMINQISNYWKKAGLVKGRKTL--G 641
QY 448 ED 449
DB 642 EN 643

RESULT 3
ID YHV5_YEAST STANDARD: PPT: 1228 AA.
AC P38851:
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE Hypothetical 143.6 kDa protein in SPO16-REC104 intergenic region.
GN YHR155W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN S298c / AK972;
RX MEDLINE-94378003; PubMed-8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favetto A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nham M., Pitvin L., Piles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vianati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2977-2982(1994).
CC 1- SIMILARITY: STRONG TO YEAST STP3 SAME TO S POWER SPAC19AR 62
CC 1- SIMILARITY: CONTAINS 1 PH DOMAIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC FMRL: 010497; AAE68977.1;
CC PIR: S46754; S46754.
CC SGD: S0001198; YHR155W
CC InterPro: IPR001849; PH.
CC Pfam: PF00169; PH; 1.
CC SMART: SM00233; PH; 1.
CC PROSITE: PS00004; PH_DOMAIN; 1.
CC Hypothetical protein.
CC FT DOMAIN 308 421 PH.
CC SEQUENCE 1228 AA; 14358 MW; C8872FAE8270A4B6 CRC64;

```

```

Query Match 3.9%; Score 110 %; DB 1; Length 1228;
Best local similarity 84%; Pred No 4;
Matches 88; Conservative 78; Mismatches 150; Indels 151; Gaps 23;
58 HSFEKKNFEVYRSKYLRSIDLYNENLANKIVVSEKAGIMAYHPCPKNIFAKAFSYLSHII 117
QY

```

```

DB 571 YPSLKVDTQPSISIF-SVNHDFELAVPKFLVILPYSYVWPNKCKAFKASMAFVITNII 629
QY 118 PEETDNCINIMKTGGDYATSETNFIKIDPQTLFLEIKVIYSKYVAVNIATSHRIVIS 177
DB 630 V-----YLNISGESVLRKID--LIDIDISDEYDK-----SEKRVSS 662
QY 178 A-----GNILNMGILSIVK-----GRKYVILFKIPSSVPEKCK-KSPFALEK- 219
DB 663 FMLHMGGHGLPFTNMSVFIDPEAVAKSLQELFENFAMHPIKPKFKEVLFPFELIPEFN 722
QY 220 ---VVCISPRS-----LQPSVYRSQNTENIVFIECPF-----KLIHVKLAIA 262
DB 723 EKLIKDNLSSEHSYKSDYDLKSYIDHFNFN-----ETPMELMSKKRIERFAW 776
QY 263 YIPGVNWAQISFHKETKTFH--FYDPKTKKEVSTKFTYDALVL-----YHH 408
DB 777 YFQD-----NEKVGSKTLFHLFGDK-----SQVFPSSFLCKKGSNNLNNSYWR 822
QY 309 INAYEEDGHVVEDI-----VAYRDNLSYDMFLKLLKDKDFE--VNNKLTSPITCKRFV 359
DB 823 IRPAKEDASQFELCPKLPQLNFTSNFIKDLWLKUNENFKLVLPQVTKIKO----- 877
QY 360 VPIQYDKDAEVGSLVKLPTSATAVKKGSIYQPEIL-----CRGIEIPR 406
DB 878 ---GYFEFVEEGHVKFP-----LCHPLLLKVKRFLIAHLLSQESLAKKD 921
QY 407 --VNYDNGKKYKYVYVATEVQWSPVPTKIATLNVOTKEVLHWGEHC 451
DB 922 LAIAYFN YVESID FIKTKYKIMJFFIHNWALNYC 959

RESULT 4
SACS_HUMAN STANDARD: PPT: 3829 AA.
ID SACS_HUMAN
AC Q9NZJ4; Q94835;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Sacsin.
GN SACS OR KIAA0730.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND VARIANT ALA-2619.
RX MEDLINE-20120709; PubMed-10655055;
RA Endert J.C., Berube P., Merleier J., Dore C., Lepaque P., Gu B.,
RA Rouchard J.-P., Mathieu J., Molancon S.R., Schalling M., Lander E.S.,
RA Morgan K., Hudson T.J., Richter A.;
RA "AFSACS, a spastic ataxia common to northwestern Quebec, is caused by
PT mutations in a new gene encoding an 11.5 kb ORF."
PL Nat. Genet. 24:120-125(2000).
RN [2]
RP SEQUENCE OF 2825-3829 FROM N.A.
RC TISSUE-Brain;
RX MEDLINE-99087487; PubMed-9872452;
RA Nagase T., Ishikawa K., I., Suyama M., Kikuno M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
PL DNA Res. 5:277-286(1998).
CC 1- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
CC 1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE CENTRAL NERVOUS
CC SYSTEM, ALSO FOUND IN SKELETAL MUSCLE AND AT LOW LEVELS IN
CC PANCREAS.
CC 1- DISEASE: DEFECTS IN SACS ARE THE CAUSE OF AUTOSOMAL RECESSIVE
CC SPASTIC ATAXIA OF CHARLEVOIX-SAGUENAY (AFSACS OR SACS). AFSACS IS
CC AN EARLY ONSET NEURODEGENERATIVE DISEASE WITH HIGH PREVALENCE IN
CC THE CHARLEVOIX-SAGUENAY-LAC-SAINT-JEAN REGION OF QUEBEC. IT IS
CC CHARACTERIZED BY AHSNI SENSORY-NEURVE CONDUCTION, REDUCED MOTOR

```



DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ran-binding protein 2 (RANBP2) (Nuclear pore complex protein Nup358)  
 DE (Nucleoporin Nup358) (358 kDa nucleoporin) (P276).  
 GN RANBP2 OR NUP358.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 GN NCBI\_TaxID=9606;  
 PN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95294031; PubMed=7775481;  
 RA Wu J., Matunis M.J., Kraemer D., Blobel G., Coutavas E.,  
 RT "Nup358, a cytoplasmically exposed nucleoporin with peptide repeats,  
 RT Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous  
 RT domain, and a leucine-rich region.";  
 RL J. Biol. Chem. 270:14209-14213(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99176415; PubMed=10078529;  
 RA Kuma K.I., Miyata T., Fukui M., Nishimoto T., Panté N., Arbi N.,  
 RT "A giant nucleoporin protein that binds Ran/Tp4.";  
 RL Nature 376:184-188(1995).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.96 ANGSTROMS) OF 1171-1304  
 RX MEDLINE=99176415; PubMed=10078529;  
 RA Vetter I.R., Nowak C., Nishimoto T., Kuhlmann J., Wittlinghofer A.,  
 RT "Structure of a Ran-binding domain complexed with a GTP  
 RT analogue: implications for nuclear transport.";  
 RL Nature 398:39-46(1999).  
 CC -!- FUNCTION: INVOLVED IN TRANSPORT FACTOR (RAN-GTP, RANP/PHRIN)-  
 CC MEDIATED PROTEIN IMPORT VIA THE F-G REPEAT-CONTAINING DOMAIN WHICH  
 CC ACTS AS A DOCKING SITE FOR SUBSTRATES. COULD ALSO HAVE ISOMERASE  
 CC OR CHAPERONE ACTIVITY AND MAY BIND RNA OR DNA. COMPONENT OF THE  
 CC NUCLEAR EXPORT PATHWAY. SPECIFIC DOCKING SITE FOR THE NUCLEAR  
 CC EXPORT FACTOR EXPORTIN-1.  
 CC -!- SUBUNIT: FORMS A TIGHT COMPLEX IN ASSOCIATION WITH RANBP1 AND THE  
 CC UBQUITIN-CONJUGATING ENZYME E2 (UBC9) (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX; CYTOPLASMIC FILAMENTS  
 CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.  
 CC -!- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.  
 CC -!- SIMILARITY: CONTAINS 8 RANBP2-TYPE ZINC FINGERS.  
 CC -!- SIMILARITY: CONTAINS 1 CYCLOPHILIN-TYPE PPIASE DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DE EMBL: L41840; AAC41758.1;  
 DE EMBL: D42063; RAA07662.1;  
 DE DB: JRP; 18-MAY-99.  
 DE Genes: HNC:9848; RANBP2.  
 DE MIM: 601181;  
 DE InterPro: IPR002130; CSA\_PPIase.  
 DE InterPro: IPR000697; RanBP1\_WASP.  
 DE InterPro: IPR000150; Ran\_RP1  
 DE InterPro: IPR001440; Tpr  
 DE InterPro: IPR001876; Znf\_RanBP  
 DE Pfam: PF00160; pro\_isomerase; 1.  
 DE Pfam: PF00515; Tpr; 1.  
 DE Pfam: PF00648; Ran\_RP1; 4.  
 DE Pfam: PF00641; zf\_RanBP; 8.  
 DE PRINTS: PR00153; CSA\_PPIASWASE.  
 DE SMART: SM00166; RanBP; 4.  
 DE SMART: SM00547; Znf\_RBP; 8.  
 DE PROSITE: PS00170; CSA\_PPIASE\_1; 1.  
 DE PROSITE: PS00072; CSA\_PPIASE\_2; 1.  
 DE PROSITE: PS00196; RANBP1; 4.

DE PROSITE, PS00196, ZF\_RANBP2\_1; 8.  
 DE PROSITE, PS00196, ZF\_RANBP2\_2; 8.  
 KW Nuclear protein, Transport, Repeat, Zinc finger, Isomerase, Rotamase;  
 KW 3D-structure.  
 FT DOMAIN 1171 1307 RANBP2-TYPE 1.  
 FT ZN\_FING 1351 1381 RANBP2-TYPE 2.  
 FT ZN\_FING 1415 1444 RANBP2-TYPE 3.  
 FT ZN\_FING 1479 1508 RANBP2-TYPE 4.  
 FT ZN\_FING 1543 1572 RANBP2-TYPE 5.  
 FT ZN\_FING 1606 1635 RANBP2-TYPE 6.  
 FT ZN\_FING 1665 1694 RANBP2-TYPE 7.  
 FT ZN\_FING 1724 1753 RANBP2-TYPE 8.  
 FT ZN\_FING 1781 1810 RANBP2-TYPE 9.  
 FT ZN\_FING 1840 1869 RANBP2-TYPE 10.  
 FT ZN\_FING 1898 1927 RANBP2-TYPE 11.  
 FT ZN\_FING 1956 1985 RANBP2-TYPE 12.  
 FT ZN\_FING 2014 2043 RANBP2-TYPE 13.  
 FT ZN\_FING 2072 2101 RANBP2-TYPE 14.  
 FT ZN\_FING 2129 2158 RANBP2-TYPE 15.  
 FT ZN\_FING 2187 2216 RANBP2-TYPE 16.  
 FT ZN\_FING 2245 2274 RANBP2-TYPE 17.  
 FT ZN\_FING 2303 2332 RANBP2-TYPE 18.  
 FT ZN\_FING 2361 2390 RANBP2-TYPE 19.  
 FT ZN\_FING 2419 2448 RANBP2-TYPE 20.  
 FT ZN\_FING 2477 2506 RANBP2-TYPE 21.  
 FT ZN\_FING 2535 2564 RANBP2-TYPE 22.  
 FT ZN\_FING 2593 2622 RANBP2-TYPE 23.  
 FT ZN\_FING 2651 2680 RANBP2-TYPE 24.  
 FT ZN\_FING 2709 2738 RANBP2-TYPE 25.  
 FT ZN\_FING 2767 2796 RANBP2-TYPE 26.  
 FT ZN\_FING 2825 2854 RANBP2-TYPE 27.  
 FT ZN\_FING 2883 2912 RANBP2-TYPE 28.  
 FT ZN\_FING 2941 2970 RANBP2-TYPE 29.  
 FT ZN\_FING 2999 3028 RANBP2-TYPE 30.  
 FT ZN\_FING 3057 3086 RANBP2-TYPE 31.  
 FT ZN\_FING 3115 3144 RANBP2-TYPE 32.  
 FT ZN\_FING 3173 3202 RANBP2-TYPE 33.  
 FT ZN\_FING 3231 3260 RANBP2-TYPE 34.  
 FT ZN\_FING 3289 3318 RANBP2-TYPE 35.  
 FT ZN\_FING 3347 3376 RANBP2-TYPE 36.  
 FT ZN\_FING 3405 3434 RANBP2-TYPE 37.  
 FT ZN\_FING 3463 3492 RANBP2-TYPE 38.  
 FT ZN\_FING 3521 3550 RANBP2-TYPE 39.  
 FT ZN\_FING 3579 3608 RANBP2-TYPE 40.  
 FT ZN\_FING 3637 3666 RANBP2-TYPE 41.  
 FT ZN\_FING 3695 3724 RANBP2-TYPE 42.  
 FT ZN\_FING 3753 3782 RANBP2-TYPE 43.  
 FT ZN\_FING 3811 3840 RANBP2-TYPE 44.  
 FT ZN\_FING 3869 3898 RANBP2-TYPE 45.  
 FT ZN\_FING 3927 3956 RANBP2-TYPE 46.  
 FT ZN\_FING 3985 4014 RANBP2-TYPE 47.  
 FT ZN\_FING 4043 4072 RANBP2-TYPE 48.  
 FT ZN\_FING 4101 4130 RANBP2-TYPE 49.  
 FT ZN\_FING 4159 4188 RANBP2-TYPE 50.  
 FT ZN\_FING 4217 4246 RANBP2-TYPE 51.  
 FT ZN\_FING 4275 4304 RANBP2-TYPE 52.  
 FT ZN\_FING 4333 4362 RANBP2-TYPE 53.  
 FT ZN\_FING 4391 4420 RANBP2-TYPE 54.  
 FT ZN\_FING 4459 4488 RANBP2-TYPE 55.  
 FT ZN\_FING 4517 4546 RANBP2-TYPE 56.  
 FT ZN\_FING 4575 4604 RANBP2-TYPE 57.  
 FT ZN\_FING 4633 4662 RANBP2-TYPE 58.  
 FT ZN\_FING 4691 4720 RANBP2-TYPE 59.  
 FT ZN\_FING 4759 4788 RANBP2-TYPE 60.  
 FT ZN\_FING 4817 4846 RANBP2-TYPE 61.  
 FT ZN\_FING 4875 4904 RANBP2-TYPE 62.  
 FT ZN\_FING 4933 4962 RANBP2-TYPE 63.  
 FT ZN\_FING 4991 5020 RANBP2-TYPE 64.  
 FT ZN\_FING 5059 5088 RANBP2-TYPE 65.  
 FT ZN\_FING 5117 5146 RANBP2-TYPE 66.  
 FT ZN\_FING 5175 5204 RANBP2-TYPE 67.  
 FT ZN\_FING 5233 5262 RANBP2-TYPE 68.  
 FT ZN\_FING 5291 5320 RANBP2-TYPE 69.  
 FT ZN\_FING 5359 5388 RANBP2-TYPE 70.  
 FT ZN\_FING 5417 5446 RANBP2-TYPE 71.  
 FT ZN\_FING 5475 5504 RANBP2-TYPE 72.  
 FT ZN\_FING 5533 5562 RANBP2-TYPE 73.  
 FT ZN\_FING 5591 5620 RANBP2-TYPE 74.  
 FT ZN\_FING 5659 5688 RANBP2-TYPE 75.  
 FT ZN\_FING 5717 5746 RANBP2-TYPE 76.  
 FT ZN\_FING 5775 5804 RANBP2-TYPE 77.  
 FT ZN\_FING 5833 5862 RANBP2-TYPE 78.  
 FT ZN\_FING 5891 5920 RANBP2-TYPE 79.  
 FT ZN\_FING 5959 5988 RANBP2-TYPE 80.  
 FT ZN\_FING 6017 6046 RANBP2-TYPE 81.  
 FT ZN\_FING 6075 6104 RANBP2-TYPE 82.  
 FT ZN\_FING 6133 6162 RANBP2-TYPE 83.  
 FT ZN\_FING 6191 6220 RANBP2-TYPE 84.  
 FT ZN\_FING 6259 6288 RANBP2-TYPE 85.  
 FT ZN\_FING 6317 6346 RANBP2-TYPE 86.  
 FT ZN\_FING 6375 6404 RANBP2-TYPE 87.  
 FT ZN\_FING 6433 6462 RANBP2-TYPE 88.  
 FT ZN\_FING 6491 6520 RANBP2-TYPE 89.  
 FT ZN\_FING 6559 6588 RANBP2-TYPE 90.  
 FT ZN\_FING 6617 6646 RANBP2-TYPE 91.  
 FT ZN\_FING 6675 6704 RANBP2-TYPE 92.  
 FT ZN\_FING 6733 6762 RANBP2-TYPE 93.  
 FT ZN\_FING 6791 6820 RANBP2-TYPE 94.  
 FT ZN\_FING 6859 6888 RANBP2-TYPE 95.  
 FT ZN\_FING 6917 6946 RANBP2-TYPE 96.  
 FT ZN\_FING 6975 7004 RANBP2-TYPE 97.  
 FT ZN\_FING 7033 7062 RANBP2-TYPE 98.  
 FT ZN\_FING 7091 7120 RANBP2-TYPE 99.  
 FT ZN\_FING 7159 7188 RANBP2-TYPE 100.  
 FT ZN\_FING 7217 7246 RANBP2-TYPE 101.  
 FT ZN\_FING 7275 7304 RANBP2-TYPE 102.  
 FT ZN\_FING 7333 7362 RANBP2-TYPE 103.  
 FT ZN\_FING 7391 7420 RANBP2-TYPE 104.  
 FT ZN\_FING 7459 7488 RANBP2-TYPE 105.  
 FT ZN\_FING 7517 7546 RANBP2-TYPE 106.  
 FT ZN\_FING 7575 7604 RANBP2-TYPE 107.  
 FT ZN\_FING 7633 7662 RANBP2-TYPE 108.  
 FT ZN\_FING 7691 7720 RANBP2-TYPE 109.  
 FT ZN\_FING 7759 7788 RANBP2-TYPE 110.  
 FT ZN\_FING 7817 7846 RANBP2-TYPE 111.  
 FT ZN\_FING 7875 7904 RANBP2-TYPE 112.  
 FT ZN\_FING 7933 7962 RANBP2-TYPE 113.  
 FT ZN\_FING 7991 8020 RANBP2-TYPE 114.  
 FT ZN\_FING 8059 8088 RANBP2-TYPE 115.  
 FT ZN\_FING 8117 8146 RANBP2-TYPE 116.  
 FT ZN\_FING 8175 8204 RANBP2-TYPE 117.  
 FT ZN\_FING 8233 8262 RANBP2-TYPE 118.  
 FT ZN\_FING 8291 8320 RANBP2-TYPE 119.  
 FT ZN\_FING 8359 8388 RANBP2-TYPE 120.  
 FT ZN\_FING 8417 8446 RANBP2-TYPE 121.  
 FT ZN\_FING 8475 8504 RANBP2-TYPE 122.  
 FT ZN\_FING 8533 8562 RANBP2-TYPE 123.  
 FT ZN\_FING 8591 8620 RANBP2-TYPE 124.  
 FT ZN\_FING 8659 8688 RANBP2-TYPE 125.  
 FT ZN\_FING 8717 8746 RANBP2-TYPE 126.  
 FT ZN\_FING 8775 8804 RANBP2-TYPE 127.  
 FT ZN\_FING 8833 8862 RANBP2-TYPE 128.  
 FT ZN\_FING 8891 8920 RANBP2-TYPE 129.  
 FT ZN\_FING 8959 8988 RANBP2-TYPE 130.  
 FT ZN\_FING 9017 9046 RANBP2-TYPE 131.  
 FT ZN\_FING 9075 9104 RANBP2-TYPE 132.  
 FT ZN\_FING 9133 9162 RANBP2-TYPE 133.  
 FT ZN\_FING 9191 9220 RANBP2-TYPE 134.  
 FT ZN\_FING 9259 9288 RANBP2-TYPE 135.  
 FT ZN\_FING 9317 9346 RANBP2-TYPE 136.  
 FT ZN\_FING 9375 9404 RANBP2-TYPE 137.  
 FT ZN\_FING 9433 9462 RANBP2-TYPE 138.  
 FT ZN\_FING 9491 9520 RANBP2-TYPE 139.  
 FT ZN\_FING 9559 9588 RANBP2-TYPE 140.  
 FT ZN\_FING 9617 9646 RANBP2-TYPE 141.  
 FT ZN\_FING 9675 9704 RANBP2-TYPE 142.  
 FT ZN\_FING 9733 9762 RANBP2-TYPE 143.  
 FT ZN\_FING 9791 9820 RANBP2-TYPE 144.  
 FT ZN\_FING 9859 9888 RANBP2-TYPE 145.  
 FT ZN\_FING 9917 9946 RANBP2-TYPE 146.  
 FT ZN\_FING 9975 10004 RANBP2-TYPE 147.  
 FT ZN\_FING 10033 10062 RANBP2-TYPE 148.  
 FT ZN\_FING 10091 10120 RANBP2-TYPE 149.  
 FT ZN\_FING 10159 10188 RANBP2-TYPE 150.  
 FT ZN\_FING 10217 10246 RANBP2-TYPE 151.  
 FT ZN\_FING 10275 10304 RANBP2-TYPE 152.  
 FT ZN\_FING 10333 10362 RANBP2-TYPE 153.  
 FT ZN\_FING 10391 10420 RANBP2-TYPE 154.  
 FT ZN\_FING 10459 10488 RANBP2-TYPE 155.  
 FT ZN\_FING 10517 10546 RANBP2-TYPE 156.  
 FT ZN\_FING 10575 10604 RANBP2-TYPE 157.  
 FT ZN\_FING 10633 10662 RANBP2-TYPE 158.  
 FT ZN\_FING 10691 10720 RANBP2-TYPE 159.  
 FT ZN\_FING 10759 10788 RANBP2-TYPE 160.  
 FT ZN\_FING 10817 10846 RANBP2-TYPE 161.  
 FT ZN\_FING 10875 10904 RANBP2-TYPE 162.  
 FT ZN\_FING 10933 10962 RANBP2-TYPE 163.  
 FT ZN\_FING 10991 11020 RANBP2-TYPE 164.  
 FT ZN\_FING 11059 11088 RANBP2-TYPE 165.  
 FT ZN\_FING 11117 11146 RANBP2-TYPE 166.  
 FT ZN\_FING 11175 11204 RANBP2-TYPE 167.  
 FT ZN\_FING 11233 11262 RANBP2-TYPE 168.  
 FT ZN\_FING 11291 11320 RANBP2-TYPE 169.  
 FT ZN\_FING 11359 11388 RANBP2-TYPE 170.  
 FT ZN\_FING 11417 11446 RANBP2-TYPE 171.  
 FT ZN\_FING 11475 11504 RANBP2-TYPE 172.  
 FT ZN\_FING 11533 11562 RANBP2-TYPE 173.  
 FT ZN\_FING 11591 11620 RANBP2-TYPE 174.  
 FT ZN\_FING 11659 11688 RANBP2-TYPE 175.  
 FT ZN\_FING 11717 11746 RANBP2-TYPE 176.  
 FT ZN\_FING 11775 11804 RANBP2-TYPE 177.  
 FT ZN\_FING 11833 11862 RANBP2-TYPE 178.  
 FT ZN\_FING 11891 11920 RANBP2-TYPE 179.  
 FT ZN\_FING 11959 11988 RANBP2-TYPE 180.  
 FT ZN\_FING 12017 12046 RANBP2-TYPE 181.  
 FT ZN\_FING 12075 12104 RANBP2-TYPE 182.  
 FT ZN\_FING 12133 12162 RANBP2-TYPE 183.  
 FT ZN\_FING 12191 12220 RANBP2-TYPE 184.  
 FT ZN\_FING 12259 12288 RANBP2-TYPE 185.  
 FT ZN\_FING 12317 12346 RANBP2-TYPE 186.  
 FT ZN\_FING 12375 12404 RANBP2-TYPE 187.  
 FT ZN\_FING 12433 12462 RANBP2-TYPE 188.  
 FT ZN\_FING 12491 12520 RANBP2-TYPE 189.  
 FT ZN\_FING 12559 12588 RANBP2-TYPE 190.  
 FT ZN\_FING 12617 12646 RANBP2-TYPE 191.  
 FT ZN\_FING 12675 12704 RANBP2-TYPE 192.  
 FT ZN\_FING 12733 12762 RANBP2-TYPE 193.  
 FT ZN\_FING 12791 12820 RANBP2-TYPE 194.  
 FT ZN\_FING 12859 12888 RANBP2-TYPE 195.  
 FT ZN\_FING 12917 12946 RANBP2-TYPE 196.  
 FT ZN\_FING 12975 13004 RANBP2-TYPE 197.  
 FT ZN\_FING 13033 13062 RANBP2-TYPE 198.  
 FT ZN\_FING 13091 13120 RANBP2-TYPE 199.  
 FT ZN\_FING 13159 13188 RANBP2-TYPE 200.  
 FT ZN\_FING 13217 13246 RANBP2-TYPE 201.  
 FT ZN\_FING 13275 13304 RANBP2-TYPE 202.  
 FT ZN\_FING 13333 13362 RANBP2-TYPE 203.  
 FT ZN\_FING 13391 13420 RANBP2-TYPE 204.  
 FT ZN\_FING 13459 13488 RANBP2-TYPE 205.  
 FT ZN\_FING 13517 13546 RANBP2-TYPE 206.  
 FT ZN\_FING 13575 13604 RANBP2-TYPE 207.  
 FT ZN\_FING 13633 13662 RANBP2-TYPE 208.  
 FT ZN\_FING 13691 13720 RANBP2-TYPE 209.  
 FT ZN\_FING 13759 13788 RANBP2-TYPE 210.  
 FT ZN\_FING 13817 13846 RANBP2-TYPE 211.  
 FT ZN\_FING 13875 13904 RANBP2-TYPE 212.  
 FT ZN\_FING 13933 13962 RANBP2-TYPE 213.  
 FT ZN\_FING 13991 14020 RANBP2-TYPE 214.  
 FT ZN\_FING 14059 14088 RANBP2-TYPE 215.  
 FT ZN\_FING 14117 14146 RANBP2-TYPE 216.  
 FT ZN\_FING 14175 14204 RANBP2-TYPE 217.  
 FT ZN\_FING 14233 14262 RANBP2-TYPE 218.  
 FT ZN\_FING 14291 14320 RANBP2-TYPE 219.  
 FT ZN\_FING 14359 14388 RANBP2-TYPE 220.  
 FT ZN\_FING 14417 14446 RANBP2-TYPE 221.  
 FT ZN\_FING 14475 14504 RANBP2-TYPE 222.  
 FT ZN\_FING 14533 14562 RANBP2-TYPE 223.  
 FT ZN\_FING 14591 14620 RANBP2-TYPE 224.  
 FT ZN\_FING 14659 14688 RANBP2-TYPE 225.  
 FT ZN\_FING 14717 14746 RANBP2-TYPE 226.  
 FT ZN\_FING 14775 14804 RANBP2-TYPE 227.  
 FT ZN\_FING 14833 14862 RANBP2-TYPE 228.  
 FT ZN\_FING 14891 14920 RANBP2-TYPE 229.  
 FT ZN\_FING 14959 14988 RANBP2-TYPE 230.  
 FT ZN\_FING 15017 15046 RANBP2-TYPE 231.  
 FT ZN\_FING 15075 15104 RANBP2-TYPE 232.  
 FT ZN\_FING 15133 15162 RANBP2-TYPE 233.  
 FT ZN\_FING 15191 15220 RANBP2-TYPE 234.  
 FT ZN\_FING 15259 15288 RANBP2-TYPE 235.  
 FT ZN\_FING 15317 15346 RANBP2-TYPE 236.  
 FT ZN\_FING 15375 15404 RANBP2-TYPE 237.  
 FT ZN\_FING 15433 15462 RANBP2-TYPE 238.  
 FT ZN\_FING 15491 15520 RANBP2-TYPE 239.  
 FT ZN\_FING 15559 15588 RANBP2-TYPE 240.  
 FT ZN\_FING 15617 15646 RANBP2-TYPE 241.  
 FT ZN\_FING 15675 15704 RANBP2-TYPE 242.  
 FT ZN\_FING 15733 15762 RANBP2-TYPE 243.  
 FT ZN\_FING 15791 15820 RANBP2-TYPE 244.  
 FT ZN\_FING 15859 15888 RANBP2-TYPE 245.  
 FT ZN\_FING 15917 15946 RANBP2-TYPE 246.  
 FT ZN\_FING 15975 16004 RANBP2-TYPE 247.  
 FT ZN\_FING 16033 16062 RANBP2-TYPE 248.  
 FT ZN\_FING 16091 16120 RANBP2-TYPE 249.  
 FT ZN\_FING 16159 16188 RANBP2-TYPE 250.  
 FT ZN\_FING 16217 16246 RANBP2-TYPE 251.  
 FT ZN\_FING 16275 16304 RANBP2-TYPE 252.  
 FT ZN\_FING 16333 16362 RANBP2-TYPE 253.  
 FT ZN\_FING 16391 16420 RANBP2-TYPE 254.  
 FT ZN\_FING 16459 16488 RANBP2-TYPE 255.  
 FT ZN\_FING 16517 16546 RANBP2-TYPE 256.  
 FT ZN\_FING 16575 16604 RANBP2-TYPE 257.  
 FT ZN\_FING 16633 16662 RANBP2-TYPE 258.  
 FT ZN\_FING 16691 16720 RANBP2-TYPE 259.  
 FT ZN\_FING 16759 16788 RANBP2-TYPE 260.  
 FT ZN\_FING 16817 16846 RANBP2-TYPE 261.  
 FT ZN\_FING 16875 16904 RANBP2-TYPE 262.  
 FT ZN\_FING 16933 16962 RANBP2-TYPE 263.  
 FT ZN\_FING 16991 17020 RANBP2-TYPE 264.  
 FT ZN\_FING 17059 17088 RANBP2-TYPE 265.  
 FT ZN\_FING 17117 17146 RANBP2-TYPE 266.  
 FT ZN\_FING 17175 17204 RANBP2-TYPE 267.  
 FT ZN\_FING 17233 17262 RANBP2-TYPE 268.  
 FT ZN\_FING 17291 17320 RANBP2-TYPE 269.  
 FT ZN\_FING 17359 17388 RANBP2-TYPE 270.  
 FT ZN\_FING 17417 17446 RANBP2-TYPE 271.  
 FT ZN\_FING 17475 17504 RANBP2-TYPE 272.  
 FT ZN\_FING 17533 17562 RANBP2-TYPE 273.  
 FT ZN\_FING 17591 17620 RANBP2-TYPE 274.  
 FT ZN\_FING 17659 17688 RANBP2-TYPE 275.  
 FT ZN\_FING 17717 17746 RANBP2-TYPE 276.  
 FT ZN\_FING 17775 17804 RANBP2-TYPE 277.  
 FT ZN\_FING 17833 17862 RANBP2-TYPE 278.  
 FT ZN\_FING 17891 17920 RANBP2-TYPE 279.  
 FT ZN\_FING 17959 17988 RANBP2-TYPE 280.  
 FT ZN\_FING 18017 18046 RANBP2-TYPE 281.  
 FT ZN\_FING 18075 18104 RANBP2-TYPE 282.  
 FT ZN\_FING 18133 18162 RANBP2-TYPE 283.  
 FT ZN\_FING 18191 18220 RANBP2-TYPE 284.  
 FT ZN\_FING 18259 18288 RANBP2-TYPE 285.  
 FT ZN\_FING 18317 18346 RANBP2-TYPE 286.  
 FT ZN\_FING 18375 18404 RANBP2-TYPE 287.  
 FT ZN\_FING 18433 18462 RANBP2-TYPE 288.  
 FT ZN\_FING 18491 18520 RANBP2-TYPE 289.  
 FT ZN\_FING 18559 18588 RANBP2-TYPE 290.  
 FT ZN\_FING 18617 18646 RANBP2-TYPE 291.  
 FT ZN\_FING 18675 18704 RANBP2-TYPE 292.  
 FT ZN\_FING 18733 18762 RANBP2-TYPE 293.  
 FT ZN\_FING 18791 18820 RANBP2-TYPE 294.  
 FT ZN\_FING 18859 18888 RANBP2-TYPE 295.  
 FT ZN\_FING 18917 18946 RANBP2-TYPE 296.  
 FT ZN\_FING 18975 19004 RANBP2-TYPE 297.  
 FT ZN\_FING 19033 19062 RANBP2-TYPE 298.  
 FT ZN\_FING 19091 19120 RANBP2-TYPE 299.  
 FT ZN\_FING 19159 19188 RANBP2-TYPE 300.  
 FT ZN\_FING 19217 19246 RANBP2-TYPE 301.  
 FT ZN\_FING 19275 19304 RANBP2-TYPE 302.  
 FT ZN\_FING 19333 19362 RANBP2-TYPE 303.  
 FT ZN\_FING 19391 19420 RANBP2-TYPE 304.  
 FT ZN\_FING 19459 19488 RANBP2-TYPE 305.  
 FT ZN\_FING 19517 19546 RANBP2-TYPE 306.  
 FT ZN\_FING 19575 19604 RANBP2-TYPE 307.  
 FT ZN\_FING 19633 19662 RANBP2-TYPE 308.  
 FT ZN\_FING 19691 19720 RANBP2-TYPE 309.  
 FT ZN\_FING 19759 19788 RANBP2-TYPE 310.  
 FT ZN\_FING 19817 19846 RANBP2-TYPE 311.  
 FT ZN\_FING 19875 19904 RANBP2-TYPE 312.  
 FT ZN\_FING 19933 19962 RANBP2-TYPE 313.  
 FT ZN\_FING 19991 20020 RANBP2-TYPE 314.  
 FT ZN\_FING 20059 20088 RANBP2-TYPE 315.  
 FT ZN\_FING 20117 20146 RANBP2-TYPE 316.  
 FT ZN\_FING 20175 20204 RANBP2-TYPE 317.  
 FT ZN\_FING 20233 20262 RANBP2-TYPE 318.  
 FT ZN\_FING 20291 20320 RANBP2-TYPE 319.  
 FT ZN\_FING 20359 20388 RANBP2-TYPE 320.  
 FT ZN\_FING 20417 20446 RANBP2-TYPE 321.  
 FT ZN\_FING 20475 20504 RANBP2-TYPE 322.  
 FT ZN\_FING 20533 20562 RANBP2-TYPE 323.  
 FT ZN\_FING 20591 20620 RANBP2-TYPE 324.  
 FT ZN\_FING 20659 20688 RANBP2-TYPE 325.  
 FT ZN\_FING 20717 20746 RANBP2-TYPE 326.  
 FT ZN\_FING 20775 20804 RANBP2-TYPE 327.  
 FT ZN\_FING 20833 20862 RANBP2-TYPE 328.  
 FT ZN\_FING 20891 20920 RANBP2-TYPE 329.  
 FT ZN\_FING 20959 20988 RANBP2-TYPE 330.  
 FT ZN\_FING 21017 21046 RANBP2-TYPE 331.  
 FT ZN\_FING 21075 21104 RANBP2-TYPE 332.  
 FT ZN\_FING 21133 21162 RANBP2-TYPE 333.  
 FT ZN\_FING 21191 21220 RANBP2-TYPE 334.  
 FT ZN\_FING 21259 21288 RANBP2-TYPE 335.  
 FT ZN\_FING 21317 21346 RANBP2-TYPE 336.  
 FT ZN\_FING 21375 21404 RANBP2-TYPE 337.  
 FT ZN\_FING 21433 21462 RANBP2-TYPE 338.  
 FT ZN\_FING 21491 21520 RANBP2-TYPE 339.  
 FT ZN\_FING 21559 21588 RANBP2-TYPE 340.  
 FT ZN\_FING 21617 21646 RANBP2-TYPE 341.  
 FT ZN\_FING



RA Grotbeck E.J., Davis N.W., Lim A., Dimelanta R.T., Potamouisis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
PL Nature 409:529-533(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-0157:H7 / RMD 0509952;  
PX MEDLINE-21156231; PubMed-11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
O157:H7 and genomic comparison with a laboratory strain K-12";  
PL DNA Res. 8:11-22(2001).  
CC -!- FUNCTION: INVOLVED IN ADHERENCE.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane  
(CC (Probable).  
CC -!- SIMILARITY: BELONGS TO THE UPF0141 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U50906; AAA66312.1; ALT\_INIT.  
DR EMBL; AE005546; AG58308.1; ALT\_INIT.  
DR EMBL; AP002564; BAB37476.1; ALT\_INIT.  
DR InterPro; IPR003371; DUF146.  
DR Pfam; PF02418; DUF146; 1.  
KW Transmembrane; Outer membrane; Signal; Complete proteome.  
FT SIGNAL 1 25  
FT CHAIN 26 541  
FT TRANSMEM 61 81  
FT TRANSMEM 111 131  
FT TRANSMEM 142 162  
FT TRANSMEM 265 285  
FT SEQUENCE 541 AA; 60639 MW; 49C140B666FA0F84 CPO64;  
SQ  
Query Match 3.8%; Score 107; DB 1; Length 541;  
Best Local Similarity 20.6%; Pred. No 2;  
Matches 80; Conservative 58; Mismatches 130; Indels 120; Gaps 19;  
QY 64 NGEVYRSKYLRSITYNCNEANRIVVSEFGTMYDPCKNIFAKA-----FSYLSHTI 117  
DB 152 SGLSFCACQFAYKDAKNKAFSPYILASRFATYT-PFNLNYPALAAKHORLLSTANTV 210  
QY 118 PEFTDNCNLINIMKTGDYYA-----TSETNFRKIDPOTLETLDK-----V 158  
DB 211 PYFQ-----LSVRDGTIDTYLVLCGESVRVDNMSLYGYTRSTPQ-VEAQRKQIKLPNQAI 265  
QY 159 DYKYVAVNL-----ATSHPHYSAGNILNM-----GTSYVD 190  
DB 266 SGAPYITALSVPLSLTADSVLSDHIDHNPYDNIINMANQAGFOTFWLSSQSAFRQNGTAVTS 325  
QY 191 -----KGPTKYVLEKIPSSVPEK-EKKKSCFKHL-----EWCSCI--PSRSLLOP 232  
DB 326 IAMPAMETVYVPGDELLPHLSAQOQNTQKKLVHLNCHSHPACSAFYQSSAVFQP 385  
QY 233 S-----YHHSFGITENYI-----VFIEQPFKIDIVKLATAYIRGVNNAAS 271  
DB 386 QDDQADACYNDSIHYTDSLLQGVPELLKDRRASVMYFADHCLPPTFK-KNVYFHGCFEAS 444  
QY 272 CLSFHKEDKTFW-----HFVDRKTKKEVSTKFTYDALVYHHINAY-----EEDCHVVF 320  
DB 445 QQAYHVPWFVIWSPVLGDGVDRITENNIFSTANNYL-----INAMGVYTKPEQQTIFR 499  
QY 321 DIVAYRNSL-----YDMFYLLKLDKDF 343

DB 500 VIVHYKDSLVVDANHLVDYVYVLRKEF 527  
RESULT 8  
ID RP94\_FOWPV STANDARD; PRT; 798 AA.  
AC Q9J589;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PNA-polymerase-associated transcription specificity factor (PAP94).  
GN FV141.  
OS Fowlpox virus (FPV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Avipoxvirus.  
OX NCBI\_TaxID:10261;  
RN [1]  
RP SEQUENCE FROM N.A.  
EX MEDLINE-29193820; PubMed-10749156.  
RA Afonso C.L., Taiman E.R., Fu Z., Zsak L., Kutish G.F., Rock D.L.;  
RT "The genome of fowlpox virus";  
PL J. Virol. 74:3815-3831(2000)  
CC -!- FUNCTION: ASSOCIATES WITH THE PNA POLYMERASE, REQUIRES FOR THE  
CC TRANSCRIPTION OF EARLY GENES. POSSIBLY MEDIATES THE BINDING OF  
CC THE CORE PNA POLYMERASE TO VETF.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF198100; AAP44485.1;  
DR InterPro; IPR004974; Pox\_Pap94.  
DR Pfam; PF01294; Pox\_Pap94; 1  
KW Transcription; Late protein.  
SQ SEQUENCE 798 AA; 95320 MW; 2A3EE5453537B03 CPO64;  
Query Match 3.8%; Score 107; DB 1; Length 748;  
Best Local Similarity 21.8%; Pred. No 4 3;  
Matches 94; Conservative 64; Mismatches 171; Indels 102; Gaps 24;  
QY 73 YLRSDT-----YNCNEANRIVVSEFGTMYDPCKNI-----FAKAF 110  
DB 18 YIKDDTVKESYODFIEKNKELFICNLNNMTIDEIKLLYITIDQIDLUKSLVAIF 77  
QY 111 SYLSHTIPE--PTUNLINI--MKPDQDYAISEINPIPKIDPQITELFKVYSYVAV 166  
DB 78 SYIGYNEKNIDHUNSSLDGDRMIGUMNVMYDTFF-----SLDFTIRQKHVI 128  
QY 167 NLATSHPHYSAGNILNMGTSIVDKGPTKY--VLEKIPSSVPEK-----EKK 211  
DB 129 -LVNDEGNDFNINYSFTTSLSYK-EDKYEQVNVNIPFNMKELLSYVSKNLDQLRFSKK 186  
QY 212 KSDPYHL--EYVYTSFSESLATSYVHSFGITERYIVFIPQYFYLIVYATAYIPQW 259  
DB 187 YLDFAYLCLRNIGIKISKPK-YNVRYIFNVVIDELTIPVIK----DYLDVKKVYLEETNK 241  
QY 270 ASCLSTFKEDK:WHEVFKIKREVSTKPYTALVYHHINAYE-----EEDCHVVF 420  
DB 242 AYPNPNFNQKPYEYFW-----GRVITPKFNPNPLYSVFETISYGTIDIMLTINPQVTF 296  
QY 321 FIVAYPANSIYDMFYLLKLDKDFEYVNNKILSIFQKPPVVPVQYK-DAEVSNSLVKLP 379  
DB 297 E---PRKNPI-EYIYVSEL-KFWEEGSGVDFVP-CEHETAIIDAKKVSLEYEYENIN 348  
QY 360 SATAVKKEKSTYQVPTFIEGIFLPPVNVYDYNKKYKYVYATVWSEVPIKIAKINVA 439  
DB 349 IAKYIYVEDGLAYCN---LC-GINIQIQLND-----ATDV-----TKISLINVT 388  
QY 440 TKEVLHWGEDH 450

```

1db 389 YNKSIFMSDEPY 399
RESULT 9
ID NAF2_DROME STANDARD: PRT; 841 AA.
AC Q9VW74;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Nuclear RNA export factor 2.
GN NAF2 OR GC4118.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Mandibulata, Pancrustacea, Hexapoda,
OC Insecta, Pterygota, Neoptera, Echteleygota, Diptera, Brachycera,
OC Muscomorpha, Ephydroidea, Drosophilidae, Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21648995; PubMed=11740633;
RA Herold A., Klymenko T., Lzaurrelle E.,
RT "NXF1/p15 heterodimers are essential for mRNA nuclear export in
RT Drosophila."
RL RNA 7:1748-1749(2001)
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkelley;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gokeyne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins P.A., Galie P.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C.A., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew P.M., Basu A., Baxendale J., Beyrataroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Brottier P.,
RA Borkova D., Botchan P.P., Bouck T., Brockstein P., Bolshakov S.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T.,
RA Cherry J.M., Crawley S., Dahle K.C., Davenport L.R., Davies P.,
RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.R., Ferrar C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Itahli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei H., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Parle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard T., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong P., Sun F.,
RA Svrtikas K., Tector C., Turner K., Venturi E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage F., Worley K.P., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavori J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.S.,
RA Gibbs K.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000)
RN [3]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dersett V., Drosnok D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller P., Li P.W., Liao G.,
RA Miranda A., Mundall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
PA Patel S., Phouanavong S., Wan K.H., Yu C., Lewis S.F., Rubin G.M.,
PA Celnik S.;
PA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May be involved in the export of mRNA from the nucleus
CC to the cytoplasm.
CC -!- SUBUNIT: Interacts with NXT1.
CC -!- SUBCELLULAR LOCATION: Localized in both the nucleoplasm and the
CC cytoplasm. Not detected at the nuclear rim.
CC -!- TISSUE SPECIFICITY: Expressed ubiquitously.
CC -!- DEVELOPMENTAL STAGE: Expressed throughout embryonic development.
CC -!- MISCELLANEOUS: THE PNA-binding domain is a new conserved PNA-type
CC domain.
CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
CC -!- SIMILARITY: CONTAINS 1 RIF2 DOMAIN
CC -!- SIMILARITY: CONTAINS 1 PNA RECOGNITION MOTIF (PPM)
CC -!- SIMILARITY: CONTAINS 1 PNA DOMAIN
CC -!- SIMILARITY: CONTAINS 1 PNA DOMAIN
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AJ312282; CAC41644.1; -
CC EMBL: AF033527; AAF49447.1; -
CC EMBL: AY075582; AAL68387.1; -
CC FlyBase: FBgn0036640; CG4118.
CC InterPro: IPR001611; LRR.
CC Pfam: PF00560; LRR; 1.
CC PROSITE: PS0177; NTF2 DOMAIN; 1.
CC Transport: mRNA transport; Nuclear protein; RNA binding; Repeat;
CC Leucine-rich repeat; Multigene family.
CC DOMAIN 325 409 PNA-BINDING (PPM) (BY SIMILARITY).
CC REPEAT 473 496 LRR 1.
CC REPEAT 498 522 LRR 2.
CC REPEAT 524 548 LRR 3.
CC REPEAT 556 579 LRR 4.
CC DOMAIN 585 758 NTF2.
CC DOMAIN 788 831 UBA-LIKE (BY SIMILARITY).
CC SEQUENCE 841 AA; 96634 MW; 581CEBUC57FF098 CRC64;
Query Match 3.8%; Score 107; DB 1; Length 841;
Best Local Similarity 20.4%; Pred. No. 4.6;
Matches 93; Conservative 64; Mismatches 122; Indels 178; Gaps 29;
QY 49 HWEDGLALLHSPTFK-----NCEVYVPSKYLPSTYVNFNFANDIVVSFFCTMAYDPQC 102
DB 324 HW-----HAFMIPDSPEFNCVFVFPPFFLPDP-----TUSNFYPC 460
QY 103 -----KNIF--AKAFSYLSHTIPEFDNLINL-MKIGD---DYATSTETNFK-- 147
DB 361 YVYINTERVFLVFN-FL-IAHV-----NN-NI-EMIPIG-G-I-FHYLMMAVSVTKQHV 416
QY 148 DP-UTFLFKRVYSKYVAVNIALSPHYISA-NTINM-ISTSVKRGKRYVFKIPSSVPE 207
DB 417 DPE--ECLQKAVSQYVQVQNMNLNLEKRPUSRECLQVVMVSLSSPKILIVYL-----SVAS 469
QY 208 KKKKPSFK-----HLEVV-----SIPSPSL 240
DB 476 PRKMTTCSFDTCHRKRVVILGAHVIGMMGT-FAVLSNNWQDSSHSLSIANFKLSLV 529
QY 231 QPSYHSPSIIENYVIFEQPEFLDIVKLAIAYI-PAVNNAS-CLSPHFE---IHTFHFDV 287
DB 530 ----LHGKLCRNY-----RIPSEYVRAV-----KRVFQLTTLTKAVD 564
QY 288 PKTK--KFSVTKFYTHALVYHHIYAYEELHVVHIVAYPIINSLVIMFKKILKRD--- 342
DB 564 LQTNPGQSLOKFLCDT-----GAYELVG-----AFLEN-----YLREPFENDEFR 603

```

```

QY 343 -----FEVNNKLTSTPTCKREHVPVLPYDKDAEVSNNLVKLPTSATAVKEK----- 388
Db 604 HNLKYYSSENFIFIL-TCNNVNV--QNHQTPKIIQLSKYNPHAPLNPKNVSKASLWJF 660
QY 389 -GSIVQCPILCEGIELPRVND-----YNGK 414
Db 661 FGCTVI-VEIL--LQLPRVTHDFHSLOTDVMHYNCK 693

RESULT 10
DP3A_HELPY
ID DP3A_HELPY STANDARD; PRT; 1211 AA
AC P56157;
DT 01-NOV-1997 (rel 35, created)
DT 01-NOV-1997 (rel. 35, last sequence update)
DT 15-JUN-2002 (rel. 41, last annotation update)
DE DNA polymerase III alpha subunit (EC 2.7.7.7).
GN DNAE OR HP1460
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision, Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210,
RN [1]
PP SEQUENCE FROM N A
PC STRAIN=26695 / ATCC 700392;
PX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavang A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann P.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson P., Khakhria H.G., Glodok A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RI Nature 388:539-547(1997).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + (DNA)(N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAL
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: AE000646; AAD08502.1;
CC TIGR: HP1460;
CC DR InterPro: IPR004013; PHP_C.
CC DR InterPro: IPR003141; PHP_N.
CC DR InterPro: IPR004805; PolC_alpha.
CC DR Pfam: PF02331; PHP_N; 1.
CC DR Pfam: PF02811; PHP_C; 1.
CC DR SMART: SM00481; POLIIIAC; 1.
CC DR TIGRfam: TIGR00594; polc; 1.
CC DR Transferrase, DNA directed DNA polymerase; DNA replication,
CC Complete proteome.
CC SEQUENCE 1211 AA; 138052 MW; FB2482E47E4EB0C 38C64.

```

```

Query Match 388 Score 107; DB 1; Length 1211;
Best Local Similarity 20.44; Pred No. 7.5;
Matches 90; Conservative 60; Mismatches 157; Indels 136; Gaps 21;

QY 46 KYNHWEDGALLHSFTKNGEVY---RSKYLSDT-----YNCNIEA 85
Db 95 KNGESYENLMELSSMAYLEGFYPPPPINKLLKEHSGIITASSACLOGEVNHLNNER 154
QY 86 NRIVSEFGTMAPDPCKNIFAKAFSYLSHIPEFTDNCNLINIKTUDUYAISENF 144
Db 155 NP-----KYGAFGY-DEAFKI---ACEYQF---IFEDVYFLEIMRHE-----ILLQPF 197
QY 145 PKIDQFLETLDKVVYKSYVAVNLTATSHPHY-----PSACNITIMATSIYVPKPTKY 196
Db 198 EQVKMSLETGLKI-----IATNDTHYTPMDAKAGEVAMCVAMGKTLNKGRIKH 248
QY 197 VLFKIPSSVPEKEK---KSCFKHLEVVCSTFESRLQ-----PSYHSHGTFENYI 245
Db 249 SVHEFYIKSPPEMAKLFADIPALENTQEIADKCVLEIDLDKDKKNEPTTPSEKFKAY 307
QY 245 VFIEQPKLIVKLATAYISGVNWSLSFHEKEKTFWFFVFPETFEFVSINKFYDALVL 305
Db 308 -----AQNECLNF-----EDDASYFAYKAEGLKE-----PLVL 336
QY 306 YHHNAVFEHSHVVFIVAVKUNSLYHMPYLLKKLIDDFEV NKLITS-----IPTC 355
Db 337 V-----PKEKHQ-QYKEHKEHIEVITNMKFFCTMILVWDFPIYA 375
QY 456 KPFVPLVLYDKDAEWSNLYKLPTSATAVKEKDSYQPELCEHIEHPVNNYNNYK 415
Db 376 KEMGIPVPGPGSAGS-LVAFALKITDIDPLKYDLLFERFLNPERISMDIDTDCQRR 434
QY 416 YKYVYATEVQWSPVPTKIANKLV 438
Db 435 PREIEYME-----KYGYNV 451

RESULT 11
MY5C_HUMAN
ID MY5C_HUMAN STANDARD; PRT; 1742 AA
AC G90X4;
ET 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
RI 15 JUN 2002 (rel. 41, Last annotation update)
DE Myosin Vc (Myosin 5c).
GN MYO5C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eularchia; Eularchia; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
EX MEDLINE=11879218; PubMed=11879218;
RA Rodriguez O.C., Cheney F.E.;
RT "Human myosin-Vc is a novel class V myosin expressed in epithelial
RT cells."
RI Cell Sci 115:991-1004(2002).
CC -!- FUNCTION: May be involved in transferrin trafficking, likely to
CC power actin-based membrane trafficking in many physiologically
CC crucial tissues.
CC -!- TISSUE SPECIFICITY: Expressed chiefly in non-neuronal tissues.
CC Particularly abundant in epithelial and glandular tissues
CC including pancreas, prostate, mammary, stomach, colon and lung.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLYCOPHYLA HEAT DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```



668 TPTVTLNKHQCPENYINVPPTNPKITWENFKIHEGWHTYDILIAKSGEGDIYCHE 727  
310 N-AYEDGHVVDIVAPONSIDYD-----FYLK 337  
728 GREYFKDYINDEI...DTIYSMKELHACSEELPPIPMVAKIAINSGYVKEVAF 744  
338 KIDPFFVNNPITSTPTCFEVPVPLAYDPAE-----VGSNIVFLPT 381  
784 PIDKEIYI-----VKRDVAGCDKLGQENGEICGCKPIVKRPPIFYNLDSINYDK 834  
382 ZAVKEKESIV...CQREI.....LCESLILPVAIANCKKYVVAIEVLSWF 429  
845 MTRKEDEPTYSTGHSILSASPYLYKLRKAF--NIDILVSHUSIFVKQKSVLWEL 892  
429 VPTK 432  
893 PKSK 896

RESULT 13

PEX\_MOUSE STANDARD, PRT; 749 AA.  
AC P70569; P97439, PRT; 749 AA.  
DT 01-NOV-1997 (rel. 35, fast sequence update)  
DT 01-NOV-1997 (rel. 35, fast sequence update)  
DT 15-JUN-2002 (rel. 41, last annotation update)  
DE Metalloendopeptidase homolog PEX (EC 3.4.24.-) (Phosphate regulating  
DE neutral endopeptidase) (X-linked hypophosphatemia protein) (RPP)  
DE (Vitamin D-resistant hypophosphatemic rickets protein).  
GN PHEX OP PEX op RPP  
OS Mus musculus (Mouse).  
CC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
CC Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.  
OX NCBI TaxID=10090,  
RN [1]  
RP SEQUENCE FROM N.A.  
FX MEDLINE=9411643, PubMed=8813412,  
RA Du L., Desbarats M., Viel J., Gioroux F.H., Cawthorn G., Ecarot B.;  
RA "Cloning of the murine Pex gene implicated in X-linked  
RT hypophosphatemia and evidence for expression in bone";  
PL Genomics 36:22-28(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9721775, PubMed=9067336,  
RA Strom T.M., Francis F., Jerez R., Rowditch A., Evans M.L.,  
RA Lehrach H., Meltinger T.;  
RT "Pex gene deletions in Gy and Hfp mice provide mouse models for  
RT X-linked hypophosphatemia";  
RL Hum. Mol. Genet. 6:165-171(1997)  
RN [3]  
RP SEQUENCE FROM N.A.  
RF STRAIN=C57BL/6J;  
RX MEDLINE=9722252, PubMed=9377527;  
RA Beck I., Soumouzu Y., Martel Y., Krishnamurthy G., Gauthier C.,  
RA Goodyer C.G., Teitelbaum S.L.;  
RT "Pex/PEX tissue distribution and evidence for a deletion in the 3'  
RT region of the Pex gene in X-linked hypophosphatemic mice";  
RL J. Clin. Invest. 99:1200-1209(1997).  
CC -1- FUNCTION: PROBABLY INVOLVED IN BONE AND DENTIN MINERALIZATION  
CC AND RENAL PHOSPHATE REABSORPTION.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).  
CC -1- TISSUE SPECIFICITY: BONE.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13.  
CC This SWISS-PROT entry is C47478. It is indexed through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL: U49908; AAC16502 1;

EMBL: U73919; AAC25962.1; -  
EMBL: U73912; AAC25964.1; -  
EMBL: U73913; AAC25965.1; -  
EMBL: U73914; AAC25966.1; -  
EMBL: U73911; AAC25963.1; -  
EMBL: U73915; AAC25967.1; -  
EMBL: U75646; AAC47750.1; -  
HSSP: P08473; IDMT.  
MEROPS: M13.091; -  
MGD: MGI:107489; Phex.  
LEI: LEI:107489; Phex; Peptidase\_M13.  
PP InterPro: IPR001230; Peptidase.  
PP InterPro: IPR001230; Peptidase.  
PFam: PF01431; Peptidase\_M13; 1.  
PRINTS: PR00786; NPPRLYSIN.  
PPSIF: PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Transmembrane;  
KW signal-anchor.  
FT DOMAIN 1 20  
FT TRANSMEM 21 37  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 38 749  
FT METAL 580 580  
FT ACT\_SITE 581 581  
FT METAL 584 584  
FT METAL 642 642  
FT ACT\_SITE 646 646  
FT CARBOHYD 71 71  
FT CARBOHYD 238 238  
FT CARBOHYD 243 243  
FT CARBOHYD 290 290  
FT CARBOHYD 301 301  
FT CARBOHYD 377 377  
FT CARBOHYD 484 484  
FT CARBOHYD 736 736  
FT CONFLICT 430 430  
SQ SEQUENCE 749 AA, 86418 MW, 888FAAFC176C19A GRC64:  
D-->V (IN REF. 3)  
Query Match 3.7%; Score 105.5; DB 1; Length 749;  
Best Local Similarity 19.8%; Pctd. No. 5;  
Matches 75; Conservative 53; Mismatches 130; Indels 121; Gaps 16;

CY 150 YSKYVAVNLATSHIYDSAGNI-----LNMGTSTIVDKGRTKYVLFKIPSSVPE-----K 208  
DB 257 YNKNISLSAMIPCHIMWLYIFKVIDIPYPPKIKQDSNVVVPVQYFKDLRLGA 356  
CY 259 KPFSSE-----FKHVVVTSLSSELLLPVYHFELEEN Y 244  
DB 357 EPRFTIANYVMPXVSPINIPPPYPMPEPSPVIGTITITIPAWYVNFIFSLPY 416  
CY 245 IV---FLEQFELLEIVLAIAYPEVSWASLSLSEFHLKRWKRVKPKKPVSTKPE--- 298  
DB 417 VVGKQFVHVHFEQEKEMKMEELDSVWATVIMLEFENI WMLAALPPFAGKKAAY 472  
CY 299 -----YDAIVLYHHKAYPELVVIVVAVVPMVY-----LMYVYFIPDFEV 445  
DB 474 LAKVYPERIMXQYVA FLPKATFSTVAVNVAGPEYLAQSLFWKRS----- 523  
CY 346 NNFISIPKPPVVPVQYKQKAVVSNIVKIPISATAVKKEKISVQPELLCEGTELP 405  
DB 524 -----AVKTEWFINPTVNAVYASINQIPFAG-----FLKPPFW----- 566  
CY 406 PV-----NLYNPKY-----KYVAIEV-----QNSDVP 430  
DB 567 FELSALALVIVHFFHGHNNQKPKYKFNENLLHWWVEVLELFFLEKPKMINKVNY 526  
CY 441 TKIAKINVOIKVLEHWGED 446  
DB 627 WPRAGINVRGKPTL--GEN 643

RESULT 14  
AR01\_PNECA









GenCore version 5.1.6  
Copyright (c) 1994 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:26:34 : Search time 27 seconds  
(without alignments)  
573.202 Million cp1 updates/sec

Title: US-10-053-192-1  
Perfect score: 2825  
Sequence: 1 METIFNKEEPEPIKAEV .....MRLDHCMTFQNDLCAETE 526

Scoring table: BLOSUM62

Gapop 10 0 , Gapext 0.5

Searched: 262574 seqs, 2942222 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A-COMR pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5A-COMR pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A-COMR pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6A-COMR pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/6A-COMR pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/6A-COMR pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1071	37.9	533	4	US-09-385-259-2
2	1071	37.9	533	4	US-09-385-259-2
3	1070	37.9	533	1	US-08-498-305A-6
4	373.5	13.2	205	4	US-09-385-259-3
5	373.5	13.2	205	4	US-09-385-259-3
6	108.5	3.8	3224	2	US-08-705-660-34
7	108.5	3.8	3224	3	US-08-989-045-34
8	106	3.8	663	4	US-08-134-078-61
9	106	3.8	680	4	US-08-144-078-95
10	100.5	3.6	509	4	US-09-134-001C-2996
11	100	3.6	1568	4	US-08-181-70F-2
12	100	3.5	1568	4	US-09-458-791-2
13	100	3.5	1568	4	US-09-459-066-2
14	98	3.5	984	1	US-08-257-073-3
15	98	3.5	984	2	US-08-184-009A-120
16	98	3.5	984	2	US-08-458-356-120
17	98	3.5	984	4	US-08-460-736-120
18	98	3.5	989	4	US-08-213-419B-2
19	98	3.5	989	4	US-08-213-419B-2
20	97.5	3.5	574	4	US-09-134-001C-3141
21	96.5	3.4	3391	2	US-08-446-855A-2
22	96.5	3.4	2391	4	US-09-150-741-2
23	96	3.4	495	4	US-08-134-001C-427A
24	94.5	3.3	3135	1	US-08-323-170B-2
25	94	3.3	3135	4	US-08-954-441-2
26	94	3.3	724	3	US-08-121-964-1
27	93.5	3.3	831	4	US-09-269-861A-8

28	93	3.3	355	4	US-09-406-545C-10
29	93	3.3	1558	2	US-08-498-305A-13
30	93	3.3	1654	4	US-08-498-305A-13
31	93	3.3	1726	2	US-08-609-049A-30
32	93	3.3	1726	4	US-09-170-996-30
33	93	3.3	2522	4	US-09-251-645-13
34	92.5	3.3	1365	6	5194600-4
35	92	3.2	993	4	US-08-836-687B-30
36	90.5	3.2	303	4	US-08-961-383-292
37	90.5	3.2	1653	2	US-08-753-125-1
38	90.5	3.2	1653	4	US-09-132-271-1
39	90.5	3.2	1653	4	US-09-142-354-22
40	89.5	3.2	528	1	US-08-160-670A-5
41	89.5	3.2	942	1	US-08-141-324-14
42	89.5	3.2	942	1	US-08-541-902-14
43	89.5	3.2	968	1	US-09-180-439-3
44	89.5	3.2	968	4	US-09-180-439-4
45	89.5	3.2	1016	4	US-09-180-439-8

ALIGNMENTS

RESULT 1

US-09-385-259-2  
: Sequence 2, Application US/09064359  
: Patent No. 6201114  
: GENERAL INFORMATION:  
: APPLICANT: Aquirre, Gustavo D.  
: APPLICANT: Aquirre, Gustavo M.  
: TITLE OF INVENTION: IDENTIFICATION OF CONFIDENTIAL STATIONARY NIGHT BUSINESS  
: TITLE OF INVENTION: IN DOGS  
: FILE PERIPHERY: 1460342481  
: CURRENT APPLICATION NUMBER: US/09-385-259  
: CURRENT FILING DATE: 1998-08-06  
: EARLIER APPLICATION NUMBER: 60/103,219  
: EARLIER FILING DATE: 1998-10-06  
: NUMBER OF SEQ ID NOS: 8  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 2  
: LENGTH: 533  
: TYPE: PRT  
: ORGANISM: Canis familiaris  
US-09-385-259-2

Query Match	37.9%	Score 1071	DB 4	Length 533
Best Local Similarity	42.2%	Fred. No. 1.5e-99		
Matches 222	Conservative 92	Mismatches 188	Indels 24	Gaps 12
QY	2	ETIFNKEEPEPIKAEVQGLPTLQGVLLRNGPCMHITGDTKYNHMPDGLLHSPF 61		
DB	17	ETV-----EELSSPLTAHTVGRIPDLTLTGLSLRCGPGLEFVSGPEPFYHLDGQALLHKF 71		
QY	62	PKNGEVYVESKYLSDTVNCNEANPIVSEFGTMAYDDPKNIFAKAFSYLSHIDEPT 121		
DB	72	FREGHTVTHRPFFITDAVVPAMTEKPIVITEPTGAPPPKNTIFSPSPFPQV--EV 129		
QY	122	DMELINMKTGRVYATSEINPEKIDPQITETLTKVYKSYVAVRLATSHPHYDSAGNI 181		
DB	130	DNALVNVVYVQVYVAVTETNTEKINTEETITETIPQNTI--NYVSVATAHPIFNUGTV 189		
QY	182	LMLCISIVKQETVYVLYVYVSSVPPFPVPSVTPVQVVSIVSISLISVSYHSPGIT 241		
DB	190	VHICGCFRNESTAYNVVTV--ELGADKIDPQISVSVVQVQVQVQVQVQVHSPGIT 246		
QY	242	ENYVIEEQPKLQVVKLATAY--TGVQVWASVLSHFHWKTHVTEKTFEVSCKYT 309		
DB	247	ENYVIEEQPKLQVVKLATAY--TGVQVWASVLSHFHWKTHVTEKTFEVSCKYT 309		
QY	302	DAALVYHRIYAEVCHVVDIVAYD--NSLYDMFYKFKLQVFP--VHNPILSIPDCK 456		
DB	407	NSFNEFRHINITYEONFPLIVIGVQVWK3EEVYVNSLYLANLKENWELVKNKAKPOPEVR 466		



Db 307 SPFNLFHHINTYEDHEILVDLCWKGFEFVNYVLYANIENWEEVKNAKAK2FEVF 366  
QY 357 FVVPLQVDFKAEVGSNVLKLP-ISAIAVKEKUGSYCQPEILCEG-----IELPKVNYD- 410  
Db 367 FVPLPLNDIK-ADTGKKNLVITPNTATAIL-SPETITWLEPEVIFSGPFAFEFFQINYOK 425  
QY 411 YNGKKYKYVATEVOWSPVTKLAKLVOTKEVLHWGEDHCWSEPIFVPSDAREDEG 470  
Db 426 YGKPYIYAYGLGNHF-VDRICKLVNKTETWVQEDSPSYSEPIFVSHDPADEDDG 484  
QY 471 VLTCTVVVSEPNKAP-ELLILDARTKELGRATVNMHLDLHGME 515  
Db 485 VLSVWVSPGQKPAYLLILNAKDLSEVARAEVINPVTPLGLF 530

RESULT 4  
US-09-385-259-3  
; Sequence 3, Application US/09/385,259  
; Patent No. 6201114  
; GENERAL INFORMATION:  
; APPLICANT: Aguirre, Gustavo D.  
; APPLICANT: Acland, Gregory M.  
; APPLICANT: Ray, Kunal  
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS  
; FILE REFERENCE: 19603/2481  
; CURRENT APPLICATION NUMBER: US/09/385,259  
; CURRENT FILING DATE: 1998-08-30  
; EARLIER APPLICATION NUMBER: 60/103,219  
; EARLIER FILING DATE: 1998-10-30  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 205  
; TYPE: PPT  
; ORGANISM: Canis familiaris  
US-09-385-259-3

Query Match 13.2%, Score 373.5, DB 4, Length 205,  
Best Local Similarity 49.3%, Pred. No. 1.1e-29;  
Matches 72, Conservative 21, Mismatches 46, Indels 7, Gaps 4;  
QY 2 ETIENRKEEPEPIKAEVGGQLPTWLGQVLLNGPMMHTIGTKYNHMFDSLALHSFT 61  
Db 17 FTV-----EELSSPLTAHVTRIGRIPLWLTGSLRCGPGLEFVSGSEPYHLFDGQALLHKFD 71  
QY 62 FKNFVYVPSKYIESTYNYCNIEANPIVVSFEFTMAYPPCKNIFAKASYLSHTIPEFT 121  
Db 72 FKEGHVYHRRFIRTDAYVVRAMTEKRIIVITEFGICAFPPCKNIFSPFSYFVGV--EVT 124  
QY 122 DNCLINIMKTGDYVYATSETNFIRK1 147  
Db 130 DNALVNVYPGEDIYACTEINFITLI 155

RESULT 5  
US-09-645-370-3  
; Sequence 3, Application US/09/645,370  
; Patent No. 6428958  
; GENERAL INFORMATION:  
; APPLICANT: Aguirre, Gustavo D.  
; APPLICANT: Acland, Gregory M.  
; APPLICANT: Ray, Kunal  
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS  
; FILE REFERENCE: 19603/2481  
; CURRENT APPLICATION NUMBER: US/09/645,370  
; CURRENT FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/485,259  
; PRIOR FILING DATE: 1998-08-30  
; PRIOR APPLICATION NUMBER: 60/103,219  
; PRIOR FILING DATE: 1998-10-30

; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 205  
; TYPE: PPT  
; ORGANISM: Canis familiaris  
US-09-645-370-3

Query Match 13.2%, Score 373.5, DB 4, Length 205,  
Best Local Similarity 49.3%, Pred. No. 1.1e-29;  
Matches 72, Conservative 21, Mismatches 46, Indels 7, Gaps 2;  
QY 2 ETIENRKEEPEPIKAEVGGQLPTWLGQVLLNGPMMHTIGTKYNHMFDSLALHSFT 61  
Db 17 FTV-----EELSSPLTAHVTRIGRIPLWLTGSLRCGPGLEFVSGSEPYHLFDGQALLHKFD 71  
QY 62 FKNFVYVPSKYIESTYNYCNIEANPIVVSFEFTMAYPPCKNIFAKASYLSHTIPEFT 121  
Db 72 FKEGHVYHRRFIRTDAYVVRAMTEKRIIVITEFGICAFPPCKNIFSPFSYFVGV--EVT 124  
QY 122 DNCLINIMKTGDYVYATSETNFIRK1 147  
Db 130 DNALVNVYPGEDIYACTEINFITLI 155

RESULT 6  
US-08-705-660-34  
; Sequence 4, Application US/08/705,660  
; Patent No. 5858683  
; GENERAL INFORMATION:  
; APPLICANT: KEESEE, SUSAN  
; APPLICANT: OBAR, ROBERT  
; APPLICANT: WU, YING-JYE  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; DETECTION OF CERVICAL CANCER  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Testa, Hurwitz & Thibault  
; STREET: 125 High St.  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER AVAILABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/705,660  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GREENHALGH, DUNCAN A.  
; REGISTRATION NUMBER: 38,678  
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3224 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-705-660-34

Query Match 3.8%, Score 108.5, DB 2, Length 3224;  
Best Local Similarity 22.1%, Pred. No. 0.7,  
Matches 63, Conservative 37, Mismatches 86, Indels 94, Gaps 15;  
QY 267 VNVASCL-----STFKEDKTFHFVDEPKT-----KKEVSTFYFTALVL 305

```

Db      576 VHWAECLQKQSGLSNLSFYDQF-----YIGRSVHYWKKVLPPLLLKIKKKNSIPEPIDLPFK 631
QY      306 YHH-----INAYEDGHVFDIVAYRDNLSYD-----MFYLKK 338
Db      642 HFHSVDIQASEIVEYBEDAHIITAILDAVNGNIEDAVTAFESIKSVVSYWNLALIFHRKA 691
QY      339 LDKDFEVNKKLT--SIPTCKR-----FVPLQYUKDAEVSGLNVLKLPSTSAIVAKFKIK 489
Db      692 --EDIE--NDALSPFEEQEEKYNLRKTRDYLIKIIDSDSNL--SVVKKLPVPLESVKEMLN 747
QY      390 SYCOPILCEG-----HELPR-----VNYDYN--CKYKYVYVATEVQWSPVP 430
Db      748 SVMGELEYEGGPLYKNGSLRNADSEIKRSTPSPTRYSLSPSKSYKSPKTPPRWAEDQ 807
QY      431 TKIAKLNQVKVEL-----HWGEDHQPSE---PIFVP 460
Db      808 NSLLKMICQGVFAIKKEMQELKLNSSNSASPHRWPTENYCPDSVP 852

RESULT 7
US-08-989-045-34
: Sequence 34, Application US/08989045
: Patent No. 6027905
: GENERAL INFORMATION:
: APPLICANT: KEESEE, SUSAN
: APPLICANT: OHAR, ROBERT
: APPLICANT: WU, YING-YE
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
: TITLE OF INVENTION: DETECTION OF CERVICAL CANCER
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Testa, Hurwitz & Thibault
: STREET: 125 High St.
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,645
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: GREENHALGH, JENNAN A
: REGISTRATION NUMBER: 38,678
: REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3224 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-989-045-34

```

```

Query Match      3.8%; Score 108.5; Db 3; Length 3224;
Best Local Similarity 22.1%; Fied. No. 0.7,
Matches 64; Conservative 37; Mismatches 86; Indels 99; Gaps 15;

QY      267 VNWASCL-----SFHKEDKTWPHFVDKRT-----KKEVSTKFTYDALVL 305
Db      576 VHWAECLQKQSGLSNLSFYDQF-----YIGRSVHYWKKVLPPLLLKIKKKNSIPEPIDLPFK 631
QY      406 YHH-----INAYEDGHVFDIVAYRDNLSYD-----MFYLKK 338

```

```

Db      632 HFHSVDIQASEIVEYBEDAHIITAILDAVNGNIEDAVTAFESIKSVVSYWNLALIFHRKA 691
QY      339 LDKDFEVNKKLT--SIPTCKR-----FVPLQYUKDAEVSGLNVLKLPSTSAIVAKFKIK 489
Db      692 --EDIE--NDALSPFEEQEEKYNLRKTRDYLIKIIDSDSNL--SVVKKLPVPLESVKEMLN 747
QY      390 SYCOPILCEG-----HELPR-----VNYDYN--CKYKYVYVATEVQWSPVP 430
Db      748 SVMGELEYEGGPLYKNGSLRNADSEIKRSTPSPTRYSLSPSKSYKSPKTPPRWAEDQ 807
QY      431 TKIAKLNQVKVEL-----HWGEDHQPSE---PIFVP 460
Db      808 NSLLKMICQGVFAIKKEMQELKLNSSNSASPHRWPTENYCPDSVP 852

RESULT 8
US-09-134-078-61
: Sequence 61, Application US/09134078
: Patent No. 6368844
: GENERAL INFORMATION:
: APPLICANT: Rykina, Edward J.
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gray Cary Ware & Freidenrich LLP
: STREET: 4365 Executive Drive, Suite 1600
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,078
: FILING DATE: 13-AUG-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/949,026
: FILING DATE: 10-OCT-1997
: APPLICATION NUMBER: 60/7056,916
: FILING DATE: 06-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/024002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 858/677-1456
: TELEFAX: 858/677-1465
: INFORMATION FOR SEQ ID NO: 61:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 663 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
US-09-134-078-61

```

```

Query Match      3.8%; Score 106; Db 4; Length 663;
Best Local Similarity 21.4%; Fied. No. 0.09;
Matches 133; Conservative 65; Mismatches 210; Indels 214; Gaps 40;

QY      30 GVLLKNGFGMITIGTKY-----NIWFD-----GLALLHSFTFNGRNVYVRSKYL 74
Db      23 GKFAINCKEFPFPGSNYYMHYKSNCMIDSVLSAPQWGIKVIPIWGFIDGFSYCRPK-- 80
QY      75 PSTYCNITFANDIVVSFECTMAYPDPYKNIPAKAFSYLSHTI---PEPTUNCLINIMKT 141
Db      81 -----NTYMHPEGVGVPEGISNA--QSGFERIDYTVAKAKELGKLVIVVNN 128
QY      132 GND-----YYATSFNTNFI--KTIIDPQTIPTIDKVDVSKYVA--VNIAISHPH 174

```





```

; FILING DATE: 10-Dec-1999
; CLASSIFICATION : Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,598
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)470-4189
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-458-791-2

```

```

Query Match      3.5%; Score 100; DB 4; Length 1568,
Best Local Similarity 19.6%; Pred. No. 1.5;
Matches 71; Conservative 58; Mismatches 137; Indels 96; Gaps 15;

QY 122 ENCLINIMKT-----GDYYATSETNFIRKIDFQILETELEKVDYKIVAVNLAIS 171
DB 632 EQCPVAVEKFSGGKPKENKUNQALQVFIKSIPEKVSILGK-----SNVIVT 683
QY 172 HPHYDSAGNI--LNMGTSLVDKG--PTKYVL-----FKIPSSVPEKEKKSCFKHLEV 221
DB 684 GANETRASNITMLKGTSTCDKQVIOVSHVLDNTHMKFSLPSS--RKEMKDVCIQPDG 741
QY 222 CS-----IPSRSLQFS-----YYHSGITENYIVFIEQFFKLDIVKLA 260
DB 742 CSSVGSLSYIALPHCSLIFPATTWISGGQNTMMGNFEDVIDNLITISHELKGNINSEY 801
QY 261 TAYIRGVNWSCLSFHKEDKTFHFVDKTKKEVSTKFTYDALVLYHHIINAYEEDGHVVF 320
DB 802 VA-----TYCGFLAPSLK---SSKVRTNVTVKLRVQDTYLDGC--- 836
QY 321 DIVAYRDNLSYDMFYLLKLDKDFEVNKNLTSIPTCKRFVVPLOYDKDAEVSNLVKLP 380
DB 837 -TLQYFEDPRFTSYVFS-----EVDTELE-----VKIKEND---NPNISKKDIE 878
QY 381 ATAVKEKGSIVCOPEILCEGIELPRVNYDNGKKYKVVATEVQWSPVPTKIAKLNQVT 440
DB 879 IFLPHGNGQINCSFENITPNQDTITTIQKTKGKTASTIANSSK---KVEVKLG 936
QY 441 KE 442
DB 937 EQ 938

```

## RESULT 13

US-09-458-066-2

```

; Sequence 2, Application US/09459066
; Patent No. 6187909
; GENERAL INFORMATION:
; APPLICANT: Spriggs, Melanie
; TITLE OF INVENTION: VIRAL ENCODED SHMAPHOPHIN PROTEIN
; TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janis C. Henry
; STREET: 51 University St
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: MS-DOS/Windows 95
; SOFTWARE: Word for Windows 95, 7.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09459,066
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,598
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)470-4189
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-459-086-2

```

Query Match 3.5%; Score 100; DB 4; Length 1568;

Best Local Similarity 19.6%; Pred. No. 1.5;

Matches 71; Conservative 58; Mismatches 137; Indels 96; Gaps 15;

```

QY 122 ENCLINIMKT-----GDYYATSETNFIRKIDFQILETELEKVDYKIVAVNLAIS 171
DB 632 EQCPVAVEKFSGGKPKENKUNQALQVFIKSIPEKVSILGK-----SNVIVT 683
QY 172 HPHYDSAGNI--LNMGTSLVDKG--PTKYVL-----FKIPSSVPEKEKKSCFKHLEV 221
DB 684 GANETRASNITMLKGTSTCDKQVIOVSHVLDNTHMKFSLPSS--RKEMKDVCIQPDG 741
QY 222 CS-----IPSRSLQFS-----YYHSGITENYIVFIEQFFKLDIVKLA 260
DB 742 CSSVGSLSYIALPHCSLIFPATTWISGGQNTMMGNFEDVIDNLITISHELKGNINSEY 801
QY 261 TAYIRGVNWSCLSFHKEDKTFHFVDKTKKEVSTKFTYDALVLYHHIINAYEEDGHVVF 320
DB 802 VA-----TYCGFLAPSLK---SSKVRTNVTVKLRVQDTYLDGC--- 836
QY 321 DIVAYRDNLSYDMFYLLKLDKDFEVNKNLTSIPTCKRFVVPLOYDKDAEVSNLVKLP 380
DB 837 -TLQYFEDPRFTSYVFS-----EVDTELE-----VKIKEND---NPNISKKDIE 878
QY 381 ATAVKEKGSIVCOPEILCEGIELPRVNYDNGKKYKVVATEVQWSPVPTKIAKLNQVT 440
DB 879 IFLPHGNGQINCSFENITPNQDTITTIQKTKGKTASTIANSSK---KVEVKLG 936
QY 441 KE 442
DB 937 EQ 938

```

## RESULT 14

US-08-257-073-3

```

; Sequence 3, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALAPIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA

```

```

1  ZIP: 10036
2  COMPUTER READABLE FORM:
3  MEDIUM TYPE: Floppy disk
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patent in Release #1 0, Version #1 30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/257,073
8  FILING DATE: 09-JUN-1994
9  CLASSIFICATION: 424
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 08/075,783
12 FILING DATE: 11-JUN-1993
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US 07/852,305
15 FILING DATE: 18-MAR-1992
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 07/672,183
18 FILING DATE: 20-MAR-1991
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Frommer, William S.
21 REGISTRATION NUMBER: 25,506
22 REFERENCE/DOCKET NUMBER: 454310-2570
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (212) 840-3333
25 TELEFAX: (212) 840-0712
26 TELEX: 425066 CORTMS
27 INFORMATION FOR SEQ ID NO: 3:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 984 amino acids
30 TYPE: amino acid
31 STRANDEDNESS: linear
32 TOPOLOGY: linear
33 MOLECULE TYPE: peptide
34 FRAGMENT TYPE: internal
35 US-08-257 073 3

```

```

Query Match      3.5%; Score 98; DB 1; Length 984;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 67; Conservative 49; Mismatches 104; Indels 130; Gaps 16;

QY 125 LINIMKTGDDY----YATSETNEIR-KIDPOTLEILD--KVDSKYVAVNLAISPHYDS 177
Db 239 LQNTGTEGKFKLVVYIKENILIKWKVYGETKDTFNKKVDVKYL----- 285
QY 178 AGNILNMGTSIVDKGPKYVLEKIPSSVPEKPKKSCPKHLEVV-SIPSPSLLQPSYVHS 237
Db 286 -----INEKE-----TPETSILIHAYKEH 304
QY 238 FGI-----TENYVIFIEQPKFLDIVKLATAYIPG-VNMASTLSFHKPDKTWHFVDPKTK 292
Db 305 NGTNLIESKNYALGSDIPEKCDIL-ASNGFLSGNFENIEKFCQAL- ---LVEKENKN 356
QY 293 EVSTKFTVDALV-LYHHINAY----EEDGHVVDIVAYRONSILYDMF----- 334
Db 457 DVCYKYLSESDIVSKFKELKRAETEDDEDDYTEYKLTESIDNLIVKMKTNENNNKSELIK 416
QY 345 -----YLKILD-----KDFEVNKKLSIPTCKRFVFLQYKIAEVS 372
Db 417 LEEVDSDSLKLELMNYCSLLKQVDITGTLDNMGNGNEMDFNNLKRLLI---VHSEENINT 473
QY 373 NLVKLPTSAVAKKDGSIYCPETLCEGIELPRVNYD-----YNCK 414
Db 474 LKNKFRNAACVCLKNVDWI-----VNKRGIVLPEIYNYDELYPNHLYNCK 518

```

RESULT 15

```

US-08-184-009-120
: Sequence 120, Application US/08184009
: Patent No. 5843975
: GENERAL INFORMATION:
: APPLICANT: Paolletti, Enzo
: APPLICANT: Tartaglia, James

```

```

Query Match      3.5%; Score 98; DB 2; Length 984;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 67; Conservative 49; Mismatches 104; Indels 130; Gaps 16;

QY 125 LINIMKTGDDY----YATSETNEIR-KIDPOTLEILD--KVDSKYVAVNLAISPHYDS 177
Db 239 LQNTGTEGKFKLVVYIKENILIKWKVYGETKDTFNKKVDVKYL----- 285
QY 178 AGNILNMGTSIVDKGPKYVLEKIPSSVPEKPKKSCPKHLEVV-SIPSPSLLQPSYVHS 237
Db 286 -----INEKE-----TPETSILIHAYKEH 304
QY 238 FGI-----TENYVIFIEQPKFLDIVKLATAYIPG-VNMASTLSFHKPDKTWHFVDPKTK 292
Db 305 NGTNLIESKNYALGSDIPEKCDIL-ASNGFLSGNFENIEKFCQAL- ---LVEKENKN 356
QY 293 EVSTKFTVDALV-LYHHINAY----EEDGHVVDIVAYRONSILYDMF----- 334
Db 357 DVCYKYLSESDIVSKFKELKRAETEDDEDDYTEYKLTESIDNLIVKMKTNENNNKSELIK 416
QY 335 -----YLKILD-----KDFEVNKKLSIPTCKRFVFLQYKIAEVS 372
Db 417 LEEVDSDSLKLELMNYCSLLKQVDITGTLDNMGNGNEMDFNNLKRLLI---VHSEENINT 473
QY 373 NLVKLPTSAVAKKDGSIYCPETLCEGIELPRVNYD-----YNCK 414
Db 474 LKNKFRNAACVCLKNVDWI-----VNKRGIVLPEIYNYDELYPNHLYNCK 518

```

```

Search completed: July 15, 2003, 09:31:03
Job time : 29 secs

```





```

QY 181 ILNMGTSLIVKGRKTKYVI FKIPSSVPEKPKKKSCPKHI EVVCSIPSPSLLQPSYHSPGI 240
DB 181 ILNMGTSLIVKGRKTKYVILPKIPSSVPEKPKKKSCPKHLEVVCSIPSKSLLQPSYHSPGI 240
QY 241 TENYIVFTEQPFKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKEVSTKPYT 300
DB 241 TENYIVFTEQPFKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKEVSTKPYI 300
QY 301 DALVLYHHINAYEEDGHVDFDI VAYRNSLYDMFYLLKKDKDFEVNKKLTSTPTCKRFVV 360
DB 301 DALVLYHHINAYEEDGHVDFDI VAYRNSLYDMFYLLKKDKDFEVNKKLTSTPTCKRFVV 360
QY 361 PLOYDKDAEVSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKTKYKYV 420
DB 361 PLOYDKDAEVSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKTKYKYV 420
QY 421 ATEVQWSPVPTKIAKLVNQTKEVLHMGEDHCWSPSEPIFVPSDAREDEGGVVLTCVWSE 480
DB 421 ATEVQWSPVPTKIAKLVNQTKEVLHMGEDHCWSPSEPIFVPSDAREDEGGVVLTCVWSE 480
QY 481 PNKAPFLLILDAKTEKELGRATVNVEMHLDLHGMFIPONDGAETE 526
DB 481 PNKAPFLLILDAKTEKELGRATVNVEMHLDLHGMFIPONDGAETE 526

```

## RESULT 2

```

US-10-053-192-4
; Sequence 4, Application US/10053192
; Publication No. US20030087336A1
; GENERAL INFORMATION:
; APPLICANT: BACHMANN, Heinrich
; APPLICANT: BRUGGER, Roland
; APPLICANT: FRIEDLEIN, Arno M
; APPLICANT: WIRTZ, Gabriele M
; APPLICANT: WOGGON, Wolf-Dietrich
; APPLICANT: WYSS, Adrian
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: BETA-BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
; FILE REFERENCE: B-H-CAROTENE 15,15'-DIOXYGENASES, ...
; CURRENT APPLICATION NUMBER: US/10/053,192
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 103382.0
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 506
; TYPE: PRT
; ORGANISM: CHICKEN
US-10-053-192-4

```

```

Query Match
Best Local Similarity 96.3%; Score 2721; DB 9; Length 506;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 EEHPEPIKAEVQGLPFWLQGVLLRNGPGMHTIGDTKYNHWFDPGLALLHSFTFKNGEVY 69
DB 1 EEHPEPIKAEVQGLPFWLQGVLLRNGPGMHTIGDTKYNHWFDPGLALLHSFTFKNGEVY 60
QY 70 RSKYLRSDTYNCNIENANPIVVSFPTMAYDPCKNIFAKAFYSLSHTIPEFTDNLINIM 129
DB 61 RSKYLRSDTYNCNIENANPIVVSFPTMAYDPCKNIFAKAFYSLSHTIPEFTDNLINIM 120
QY 140 KTGDYATSETNFIKRIIPQTLTLDKVDYSKYVAVNATLSHPHYDSAGNILLNGTSLV 189
DB 121 KTGDYATSETNFIKRIIPQTLTLDKVDYSKYVAVNATLSHPHYDSAGNILLNGTSLV 180
QY 190 DKGTKYVILFKIPSSVPEKPKKKSCPKHLEVVCSIPSRLLQPSYHSPGITENYIVFTE 249
DB 181 DKGTKYVILFKIPSSVPEKPKKKSCPKHLEVVCSIPSRLLQPSYHSPGITENYIVFTE 240

```

```

QY 250 QPPKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKPVSTKPYTHAI VIYHHI 409
DB 241 QPPKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKPVSTKPYTHAI VIYHHI 400
QY 310 NAYEEDGHVDFDI VAYRNSLYDMFYLLKKDKDFEVNKKLTSTPTCKRFVVPIQYIKDAE 469
DB 301 NAYEEDGHVDFDI VAYRNSLYDMFYLLKKDKDFEVNKKLTSTPTCKRFVVPIQYIKDAE 460
QY 370 VGSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKTKYKYVYATEVQWSPV 429
DB 361 VGSNLVKLPTSATAVKEKDGSIYQPEILCEGIELPRVNDYNGKTKYKYVYATEVQWSPV 420
QY 430 PTKIAKLVNQTKEVLHMGEDHCWSPSEPIFVPSDAREDEGGVVLTCVWSEPNKAFIILL 489
DB 421 PTKIAKLVNQTKEVLHMGEDHCWSPSEPIFVPSDAREDEGGVVLTCVWSEPNKAFIILL 480
QY 490 LDKTEKELGRATVNVEMHLDLHGMF 515
DB 481 LDKTEKELGRATVNVEMHLDLHGMF 506

```

## RESULT 3

```

US-10-053-192-5
; Sequence 5, Application US/10053192
; Publication No. US20030087336A1
; GENERAL INFORMATION:
; APPLICANT: BACHMANN, Heinrich
; APPLICANT: BRUGGER, Roland
; APPLICANT: FRIEDLEIN, Arno M
; APPLICANT: WIRTZ, Gabriele M
; APPLICANT: WOGGON, Wolf-Dietrich
; APPLICANT: WYSS, Adrian
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: BETA-BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID
; FILE REFERENCE: B-H-CAROTENE 15,15'-DIOXYGENASES, ...
; CURRENT APPLICATION NUMBER: US/10/053,192
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 103382.0
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 529
; TYPE: PRT
; ORGANISM: BOVINE
US-10-053-192-5

```

```

Query Match
Best Local Similarity 35.9%; Score 1014.5; DB 9; Length 529;
Matches 226; Conservative 9; Mismatches 184; Indels 44; Gaps 20;
QY 10 EEHPEPIKAEVQGLPFWLQGVLLRNGPGMHTIGDTKYNHWFDPGLALLHSFTFKNGEV 67
DB 1 EELSSPLTAHVGTGRIPLWLTGSLRKTGPGLEFVQSEPPHYHFDGALLHRTPEKREHV 60
QY 68 YRSKYLRSDTYNCNIENANPIVVSFPTMAYDPCKNIFAKAFYSLSHTIPEFTDNL 125
DB 61 TYHPRFIRTDYAVRAMTEKPIVITREPGTTCAPPDPCKNIFSPFSPYSGV-EVDNAL 118
QY 126 INIMKTDDYATSETNFIKRIIPQTLTLDKVDYSKYVAVNATLSHPHYDSAGNILL 183
DB 119 VNYVPGEDYATSETNFIKRIIPQTLTLDKVDYSKYVAVNATLSHPHYDSAGNILL 178
QY 184 MGTSLIVKGRKTKYVILPKIPSSVPEKPKKKSCPKHLEVVCSIPSKSLLQPSYHSPGI 243
DB 179 IGCNCKNFSIANIVKIPPLQADKEPLSKETS-EIVVQFPCSRPFKYSVHSPGILIN 247
QY 244 YIVFTEQPFKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKEVSTKPYI 400
DB 238 YIVFTEQPFKLDIVKLATAYIRGVNWSCLSPHKEKDKTWFHFVDRKTKKEVSTKPYI 297
QY 301 DALVLYHHINAYEEDGHVDFDI VAYRNSLYDMFYLLKKDKDFEVNKKLTSTPTCKRFVV 454

```

```

Db      298 SPFNLFPHINTYEDNGFLVDLCWKGFEEVYNYFTLYLANRENWEEYKKNARKAPOPE 357
QY      355 CKRFVVPLOYDKDAEVSNLVLP--TSATAV--KEKDGSIYCOPEILCBG---JPLPRV 407
Db      358 VRYVLPNLNDK-ADTKNLVLTLPNTATAILCSDEFTTWLEPFVLFSSPQAEEFPOI 416
QY      408 NYD-YNGKKYKYVYATEVQHSVPPTKIAKLNVOTKEVLH--WGEDHCWPSSEDFVPSDPA 464
Db      417 NYQYCKPYTYAYGLGLNHF-VFDRICKLNVRKTKFTWFTVMQRPDSYSPSEDFVSHDPA 475
QY      465 REDEGVVLTGVVSEPNKAF-ELLILDAKTEKELGRA--TVNVEMHLDHGMF 515
Db      476 LEEDGGVLSVSVSPGAGOKPAYLLILNAKOLSEVARAEFTVEINIPVTFHGLP 529

```

## RESULT 4

```

US-09-758-269-18
; Sequence 18, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:

```

```

; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-18

```

## Query Match

```

Best Local Similarity 11.08; Score 311; DB 10; Length 538;
Matches 136; Conservative 78; Mismatches 213; Indels 144; Gaps 25;

```

```

QY      9 KEHPEPIK-AEVGGQLPTWGLVLLRNGSGMHTIGDTKYN-----HWF-DGLALLHSPTF 62
Db      56 RDETP-PVKDLPVHGFLPECLNGEFVRVGP-----NPKFDVAGVYHWFDSGDMHGVPI 108
QY      63 KNGEYVYSKYLRSDTYNCNIEANPIVVSFFGTMAYPDCKNIFAKAFSYLSHTIPEFTD 122
Db      109 KDGKATVYSRY-----VKTSLKQEEF-----FGAAKFMKIGDLKGFG 147
QY      123 NCLINI-----MKTGDYYATSETN-----FIKIDPQTLFILD 156
Db      148 LLWNVNQLRTKILNDTNGTANTALVYHGGKLLALQEAADKPVYIKVLEDGDLQTLG 207
QY      157 KVDYSKYVAVNLTATSHPHYDS-ACNLIINMGTSIVDKGRTRYVLFKIPSSVPEKEKKSCF 215
Db      208 IIDYDKRLTHSF-TAHPKVPDVTGEMFTFGYS-----HTPPYLTFRVIS-----KUGI 254
QY      216 KHLWVCSIPSRLQPSYHSFGITENYIVTFEGFKL---DIVKLATAYIKGVNNAWSC 272
Db      255 MHDVPVITIS-----EPIMHDFAITETVAIFMDLPMHFRPKEMVK-----EKKMI 300
QY      273 LSEHKEDKTFHFVDRKTKKEVSTK-FYTDALVLYHHINAYEEDGHVVFIVAYRDNLSLY 331
Db      301 YSPDPTKARFGLPRYAKDELIRWFLPNCDFIFHNANAWEEDEVVL-ITCRLENPDL 359
QY      332 DMFYLLKLDKDFEVNNKLTISPTCYFPVPLQYDKDAEVSGLVLPKTSATAVKKDGSI 391
Db      360 DMVSGKVKELNFGNELIYEM---RF-----NKKTS 389

```

```

QY      352 YQDFEILCEGLFEPVNVYTNKKYKYVYATEVQHSVPPTKIAKLNVOT 440
Db      390 -SOKKLSASAVDPRIECYTGKKQRYVVGITLDSIAKVTGIIKFDLHAEATGKRMLEV 448
QY      441 ---KEVLFHWFETHTWPSFTFVPSPTAFEFQVVI-TVVV-EPNFAPELITLAFTEF 496
Db      449 GGNITKGIYDGFGR-YGSFAIYVPPETA-EDDGYLIFVPHNENTGKS-CVIVITAKIMS 505
QY      497 ELGRATVNV--EMHLDHGMFIFQNDLGAET 525
Db      506 AEPVAVVELPHRVPGFHALFVTEBQLOBT 536

```

## RESULT 5

```

US-09-758-269-8
; Sequence 8, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:

```

```

; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-758-269-8

```

## Query Match

```

Best Local Similarity 11.08; Score 310; DB 10; Length 538;
Matches 135; Conservative 79; Mismatches 213; Indels 144; Gaps 25;

```

```

QY      9 KEHPEPIK-AEVGGQLPTWGLVLLRNGSGMHTIGDTKYN-----HWF-DGLALLHSPTF 62
Db      56 RDETP-PVKDLPVHGFLPECLNGEFVRVGP-----NPKFDVAGVYHWFDSGDMHGVPI 108
QY      63 KNGEYVYSKYLRSDTYNCNIEANPIVVSFFGTMAYPDCKNIFAKAFSYLSHTIPEFTD 122
Db      109 KDGKATVYSRY-----VKTSLKQEEF-----FGAAKFMKIGDLKGFG 147
QY      123 NCLINI-----MKTGDYYATSETN-----FIKIDPQTLFILD 156
Db      148 LLWNVNQLRTKILNDTNGTANTALVYHGGKLLALQEAADKPVYIKVLEDGDLQTLG 207
QY      157 KVDYSKYVAVNLTATSHPHYDS-ACNLIINMGTSIVDKGRTRYVLFKIPSSVPEKEKKSCF 215
Db      208 IIDYDKRLTHSF-TAHPKVPDVTGEMFTFGYS-----HTPPYLTFRVIS-----KUGI 254
QY      216 KHLWVCSIPSRLQPSYHSFGITENYIVTFEGFKL---DIVKLATAYIKGVNNAWSC 272
Db      255 MHDVPVITIS-----EPIMHDFAITETVAIFMDLPMHFRPKEMVK-----EKKMI 300
QY      273 LSEHKEDKTFHFVDRKTKKEVSTK-FYTDALVLYHHINAYEEDGHVVFIVAYRDNLSLY 441
Db      301 YSPDPTKARFGLPRYAKDELIRWFLPNCDFIFHNANAWEEDEVVL-ITCRLENPDL 459
QY      332 DMFYLLKLDKDFEVNNKLTISPTCYFPVPLQYDKDAEVSGLVLPKTSATAVKKDGSI 491
Db      360 DMVSGKVKELNFGNELIYEM---RF-----NKKTS 489
QY      392 YQDFEILCEGLFEPVNVYTNKKYKYVYATEVQHSVPPTKIAKLNVOT 440
Db      490 -SOKKLSASAVDPRIECYTGKKQRYVVGITLDSIAKVTGIIKFDLHAEATGKRMLEV 448

```







Db 477 EVGMVNRLLGSRYSYAVLAVAPWPKESGFARQDSTGELTKFYEGCH-FCGRPCFVP 535  
QY 461 SPDA-----REDEGVLTGVVSEPNKAPFLILLDAKTFKELGRATVNVEMHLLDLHGMEI 516  
Db 536 MDPAARHPRGDDGVLT-FVHDERAGTSBLLVYNAADIRLEATVQLPSRVPRFGHGTPI 594  
QY 517 PONDIGAE 524  
Db 595 FQLELEAQ 602

RESULT 12  
US-09-758-269-4  
; Sequence 4, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR FILING DATE: 2001-01-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-758-269-4

Query Match 8.3%; Score 233.5; DB 10; Length 595;  
Best Local Similarity 21.6%; Pred. No. 2e-10;  
Matches 119; Conservative 86; Mismatches 210; Indels 135; Gaps 24;  
QY 20 VQGLPTWLOGVLRNGPGMHTIGTKYHWFQGLALLHSTFKNGEYVRSYLRSDTY 79  
Db 130 IHTLPLSLNGAYTRNGPNQFPLRGPY-HLFDGDMHLAIKTHNGKATLCSYVK--TY 186  
QY 80 NCNIE-----ANRWSEFGTMAYPOCKNIFAKATS 111  
Db 187 KYNVEKOTGAPVMPNVESGNCVTASVARGALTAARVLTQY-----NPVNGI----- 234  
QY 112 YLSHTIPEFTDNCINIMKTGDYATSETNF---IPKINPQITFLDKVDYKYVAVNL 168  
CL 235 GLANTSTAFFSNRL-----FALGESDLFYAVRLTESGIEIGRYDEGSLKLSM 284  
QY 169 ATSHPHYDS-AGNILNMCTSIVDKGRTKYVLFKIPSSVPEKEKKKSCFKHLEVVCISPSR 227  
Db 285 -TAHPTDPTIGETFAFYRGVPPLT-YFEP---DSAGKOR-----DVPIF 327  
QY 228 SLQPSYVHSFTEIYIVF--IEQPFKLDIVKLATAYIRGVNWSCLSHKBDKTIWPH 285  
Db 328 SMTSPFLHDAITKRAIFAELIQLGRMNMLDIV-----LEGSPVGTDCNKTPLGV 381  
QY 286 VDRKTKKEVSTK-FYTDALVLYHINAY-PEDGHVVDIVAYRDNLSYDMFYLLKLDKDF 343  
Db 382 IPKYAGDESEMKWFEVPGFNIIHAINAWDEDDGN---SVVLIAPNIMSTIEHTLERMD--- 435  
QY 344 EVNNKLTISPTCKRFVVPLOVDKDAEVSNLVKLPTSATAVKEKDGSIYQCPILCEGIE 403  
Db 436 -----LVIALVEKVKLDIVTGIVRRHP-ISARNLD 464  
QY 404 LPRVNYNGKKYVYVATEVQVSPVPTKIATLNVQ-----TKVLHMGEDHCWPSFP 456  
Db 465 FAVINPAFLGRCSYVYVAAIGDPMPKISGVVVKLVNSKGRDNDCTVAPPMYGSQ-CYCGRP 523

QY 457 IFV--PSPDAREDEGVLTGVVSEPNKAPFLILLDAKTFKELGRATVNVEMHLL--DL 511  
Db 534 FVVARDFGNFAEEDDGVVYVHDEVTVGESKF-LVMCAKSTELFIVAVALPFRVYGH 582  
QY 512 HGMETPOND 521  
Db 583 HGLEVKESDI 592

RESULT 13  
US-09-758-269-12  
; Sequence 12, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR FILING DATE: 2001-01-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 612  
; TYPE: PRT  
; ORGANISM: Vigna unguiculata  
US-09-758-269-12

Query Match 7.8%; Score 220; DB 10; Length 612;  
Best Local Similarity 21.9%; Pred. No. 2.5e-09;  
Matches 126; Conservative 98; Mismatches 215; Indels 136; Gaps 30;  
QY 1 METINRKNKESHPPIKAE-----VQGLPTWLOGVLLKNG-- 36  
Db 116 VETALVSHEPKHPIKTAADPRVQIAGNFAVPYPAHAAQGLPVVVKIPKCIDGVYVRNGAN 175  
QY 37 PMHTIGTKYHWFQGLALLHSTFKNGEYVRSYLPSTYCNINIANPIVVSFGTM 96  
Db 176 PLYPEVAG---HFEEDGVMHVKFTNGAASACFP--TFTQMSQ-----KSLGRP 224  
QY 97 AYDPCKNI-----FAKAFSVLSHTIPEFTDNCIL-INIMKTGDYV-----ATSETNF-- 143  
Db 225 VPKAIGELHSHSGIARILLPYACIPCLVWCGMGCVANAGLVYFNNHLLAMSEDLVY 284  
QY 144 -IPKIDP-JLETLKVIYSKYVAVNIAISHPHYISA-INTINM-TSTIVIKRKYVLFK 201  
Db 285 HVPITPNDDITVGPYDFNGINSTM-IAPKILPVPVGHAI SYNVILKPKLYKFRS- 342  
QY 202 FSSV PEKCKKSCFKHLEVVCISIPSLQSYHSEFCHENYIV-IEQF--FELIV 257  
Db 343 PDGKSPDVE-----IP---LKEPTMHDFAITENFVVVPCQVVKL--- 362  
QY 258 KIATAYTEGVNWSCLSHFKELKTWFEVLFKTKKEVSTREYIDA--LVLYHHINAYEE- 314  
Db 383 ---TEMITG---GSPVVDKNTSPFGIL-HKNAKDANAMRWIDAPICPCFHLWNAWEP 435  
QY 315 --DGHVVDIVAYRDNLSYDMFYLLKLDKDFVNNKITSPTKPFVPLQYDKDAEVS 372  
Db 436 ETEEVVIGSMTPADSIFN-----EGEESLSVL-----SEIRL 470  
QY 373 NLVKLPTSATAVKEKDGSIYQCPILCF---CIELPVNYDNGKKYKYVATEVQVSP 428  
Db 471 NL-----PTKSTSTPPP-IISDAQGVNLEAGMNPKNLGRFTGFAYLALAEPP 517  
QY 429 VPTKIATLNVQKEV--LHWTFCHWPSFTFVESHPPEFEGVILCVVVSFPAPFP 486  
Db 518 KVSCHAKVLLSGHVKYMYGCEKA-FGSAFLI--FWGQKEDDYILAEVHHEKWKSE- 573





QY	+	495	FKELGPRATVNVEMHLDLHGMEFIPQNDLGAE	524
	+ + + + +			
DB	+	804	EK-----PVKPENSTD-NGLMNPGEVCGSD	827

Search completed: July 15, 2003. 09:39:20  
Job time : 57 secs

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on July 15, 2003, 09:36:25, Search time: 04 seconds  
(without alignments)  
1290 247 Million cell updates/sec

Title: US-10-053-192-1

Perfect score: 526

Sequence: 1 METIFNRKKEHPPIRAEV... MHLFQHMIFINIIAAETE 526

Scoring table: Q150

Gapop 60 0, Gapext 60 0

Searched: 671580 seqs, 205047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_protist.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteria.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	526	100.0	526	13	Q91994
2	25	4.8	547	4	Q9HAY6
3	25	4.8	547	4	Q9HVS
4	25	4.8	566	11	Q9HVS6
5	25	4.8	566	11	Q9ERN9
6	25	4.8	566	11	Q91X15
7	13	2.5	516	13	Q90WH4
8	11	2.1	420	5	Q9KKW9
9	11	2.1	620	5	Q9VESC
10	10	1.9	533	13	Q9VXC2
11	10	1.9	549	13	Q90WH3
12	8	1.5	164	2	Q9A834
13	8	1.5	287	14	Q9PFA2
14	8	1.5	293	14	Q9CTP3
15	8	1.5	333	5	Q95RP5
16	8	1.5	483	11	Q8VHP2

17	8	1.5	532	11	Q99NFI
18	8	1.5	533	4	Q16518
19	8	1.5	533	5	Q28175
20	8	1.5	533	6	Q9TVB8
21	8	1.5	533	6	Q97623
22	8	1.5	533	6	Q9XT71
23	8	1.5	533	6	Q05661
24	8	1.5	533	11	Q9VX76
25	8	1.5	533	11	Q91Z05
26	8	1.5	533	13	Q9Y175
27	8	1.5	539	4	Q96JY5
28	8	1.5	545	4	Q96J08
29	8	1.5	556	4	Q9BYV7
30	8	1.5	564	14	Q9K058
31	8	1.5	580	2	Q9JPJ5
32	8	1.5	580	16	Q93Y22
33	8	1.5	580	15	Q93Q01
34	8	1.5	724	10	Q9LKP8
35	8	1.5	856	11	Q62121
36	8	1.5	876	11	Q60752
37	8	1.5	1011	5	Q9Y148
38	8	1.5	1059	5	Q9VXK6
39	7	1.3	53	3	Q8TGT9
40	7	1.3	78	16	Q92JB4
41	7	1.3	92	16	Q8R9N7
42	7	1.3	105	2	Q54742
43	7	1.3	111	17	Q9Y8Y6
44	7	1.3	124	2	Q05206
45	7	1.3	124	10	Q49708

## ALIGNMENTS

## RESULT 1

Q91993	PRELIMINARY: PRT; 526 AA;
DI	Q91993
AC	Q91993: 01-OCT-2000 (JREMHREL, 15, Created)
LI	01-OCT-2000 (JREMHREL, 15, Last sequence update)
PT	01-DEC-2001 (JREMHREL, 19, Last annotation update)
DE	Beta carotene 15,15'-dioxygenase (EC 1.13.11.21).
GN	BCDO.
OS	Gallus gallus (Chicken).
CC	Eukaryota; Metazoa; Chordata; Craciata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
KC	STRAIN=LSL LOHMANN; TISSUE=DUODENUM;
PX	MEDLINE=20251261; PubMed=14799297;
RA	Wyss A., Wirtz G.M., Woggon W.D., Brugger R., Wyss M., Friedlein A.,
RA	Bachmann H., Hunziker W.:
FT	"Cloning and expression of beta,beta carotene-15,15'-dioxygenase";
KL	Biochem Biophys. Res Commun. 271:344-346(2000).
RN	[2]
PP	SEQUENCE FROM N.A.
KC	STRAIN=LSL LOHMANN; TISSUE=DUODENUM;
PX	MEDLINE=21134366; PubMed=11237856;
PA	Wyss A., Wirtz G.M., Woggon W.D., Brugger P., Wyss M., Friedlein A.,
RA	Kiss G., Bachmann H., Hunziker W.:
FT	"Expression pattern and localization of beta,beta-carotene 15,15'-
FT	dioxygenase in different tissues";
FL	Biochem. J. 354:531-539(2001).
DR	EMBL, AJ271386, CAB92827.1;
DR	InterPro, IPR004294; PPF65;
DR	InterPro, IPR0055; PPF65; 1;
FW	Dioxygenase, oxidoreductase.
SC	SEQUENCE 526 AA; 67484 MW; K007FAAP75FE3458 CRC64;

Query Match: 100.0%, Score 526, LR 13, Length 526;  
Best Local Similarity: 100.0%; Pred No. 0;

```

Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METFNKNKEHPPEIKAEVQGLPTWLCQVLLRNGPGMHTIGDTKYNHWFGLALLHSF 60
DB 1 METFNKNKEHPPEIKAEVQGLPTWLCQVLLRNGPGMHTIGDTKYNHWFGLALLHSF 60

QY 61 TFKNCEVYRSKYLKSDTYNCFANKIVVSEFGIMAYDPCKNIFAKAFSYLSHTIPDF 120
DB 61 TFKNCEVYRSKYLKSDTYNCFANKIVVSEFGIMAYDPCKNIFAKAFSYLSHTIPDF 120

QY 121 TDNCLINIMKTGDYYATSETNFIKIDPOTLETLDKDYSKYVAVNLTATSHPHYDSAGN 180
DB 121 TDNCLINIMKTGDYYATSETNFIKIDPOTLETLDKDYSKYVAVNLTATSHPHYDSAGN 180

QY 181 ILNMGTSIVDKGRTKYVLFKIPSSVPEKPKKSCFKHLEVCVCSIPSRSLLOPSYHSFGI 240
DB 181 ILNMGTSIVDKGRTKYVLFKIPSSVPEKPKKSCFKHLEVCVCSIPSRSLLOPSYHSFGI 240

QY 241 TENYIVFIEQPFKLDIVKLATAYIRGVNMAASCLSPHKEDKTFHFVDRKTKKEVSTKFTY 300
DB 241 TENYIVFIEQPFKLDIVKLATAYIRGVNMAASCLSPHKEDKTFHFVDRKTKKEVSTKFTY 300

QY 301 DALVLYHHINAYEDGHVDFDIWAYRDNLSYDMFYLLKLDKDFEVNKNKLTSTPTCKPFV 360
DB 301 DALVLYHHINAYEDGHVDFDIWAYRDNLSYDMFYLLKLDKDFEVNKNKLTSTPTCKPFV 360

QY 461 PLOYDKDAEFGSNI VKLPTSATAVFDFDYSIYQPFILFEGFIPFPVNYDYNCKKYKYYV 420
DB 461 PLOYDKDAEFGSNI VKLPTSATAVFDFDYSIYQPFILFEGFIPFPVNYDYNCKKYKYYV 420

QY 421 ATEVQWSPVTKIAKLVQTKVELVHWGEDHCWSPSEFIFVPSDPAREDFGVVLTCTVWVSE 480
DB 421 ATEVQWSPVTKIAKLVQTKVELVHWGEDHCWSPSEFIFVPSDPAREDFGVVLTCTVWVSE 480

QY 481 PNKAPFILLIATKTEKELSPATVNVEMHLLDRLGMEIPQNDLGAETE 526
DB 481 PNKAPFILLIATKTEKELSPATVNVEMHLLDRLGMEIPQNDLGAETE 526

RESULT 2
Q9HAY6
ID Q9HAY6 PRELIMINARY: PRT: 547 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).
GN BCDL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan W., Jang G.-P., Haeseleer F., Esumi N., Chang J., Kerrigan M.,
RA Campochiaro M., Campochiaro P., Paleczewski K., Zack D.J.;
RT "Cloning and characterization of a human beta, beta-carotene 15,15'-
RT dioxygenase that is highly expressed in the retinal pigment
RT epithelium."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases
RL EMBL: AF294900; AAG15380.1;
DR InterPro: IPR004294; RPE65.
DR Pfam: PF04055; RPE65; 1.
KW Dioxxygenase; Oxidoreductase.
SQ SEQUENCE: 547 AA; 62637 MW; F94BC8B01056F9CB CRC64;

Query Match 4.8%; Score 25; DB 4; Length 547;
Best Local Similarity 100.0%; Pred. No. 20-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 NIEANRIVVSEFGTMAPYDPCKNIF 106
DB 82 NIEANRIVVSEFGTMAPYDPCKNIF 106

RESULT 4
Q9JJS6
ID Q9JJS6 PRELIMINARY: PRT: 566 AA.
DT 01-MAR-2001 (TrEMBLrel. 15, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).
GN BCDL OR BCDL OR BCDL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wyss A., Witz G.M., Woggon W.D., Bruggen K., Wyss M., Friedlein A.,
RA Bachmann H., Hunziker W.;
RT "Expression pattern and localization of beta, beta-carotene 15,15'-
RT dioxygenase in different tissues."
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cunningham F.X.;
RA Cunningham F.X.;
RT "Identification, expression and substrate specificity of a mammalian
RT beta-carotene 15,15'-dioxygenase."
RL J. Biol. Chem. 0:0-0(2001);
DE EMBL: AJ278364; CAB92521.2;
DR EMBL: AF271298; AAG33982.1;
DR MGI: 1926923; Bcdol.
DR InterPro: IPR004294; RPE65.
DR Pfam: PF03055; RPE65; 1.

```

KW Dioxigenase; Oxidoreductase.  
SQ SEQUENCE 566 AA: 63864 MW: 184367815247892 37674:

Query Match 4.8%, Score 25, DB 11, Length 566,  
Best Local Similarity 100.0%, Pred. No. 2 le-18;  
Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106  
|||||  
DB 82 NIEANRIVVSEFGTMAYDPCKNIF 106

## RESULT 5

Q9ERN9  
ID Q9ERN9 PRELIMINARY: PRT: 566 AA  
AC Q9ERN9  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Beta, beta-carotene 15,15'-dioxigenase (EC 1.13.11.21).  
GN BCD01 OR RCD0.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=KIDNEY.  
PA Yan W., Jiang G.-F., Haseolov F., Esimi N., Chang J., Korrigan M.,  
RA Campochiaro M., Campochiaro P., Palczewski K., Zack D.J.;  
PT "Cloning and characterization of a human beta, beta carotene 15,15'-  
RT dioxigenase that is highly expressed in the retinal pigment  
RL epithelium."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases  
DR EMBL: AF294899, AAC15381.1, -  
DK MGD: MGI:1926923, Bcd01.  
DR InterPro: IPR004294; RPE65.  
DR Pfam: PF03055; RPE65, 1.  
KW Dioxigenase; Oxidoreductase.  
SQ SEQUENCE 566 AA: 63862 MW: 184367815247892 37674:

Query Match 4.8%, Score 25, DB 11, Length 566,  
Best Local Similarity 100.0%, Pred. No. 2 le-18;  
Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106  
|||||  
DB 82 NIEANRIVVSEFGTMAYDPCKNIF 106

## RESULT 6

Q91XT5  
ID Q91XT5 PRELIMINARY: PRT: 566 AA  
AC Q91XT5  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Beta-carotene 15,15'-dioxigenase.  
OS Pampus norvegicus (Pam)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=INTESTINE.  
RA Takitani K., Ban R., Tamai H.;  
PT "Regulation of beta-carotene 15,15'-dioxigenase in oxidative stress".  
PL Submitted (JUN 2001) to the EMBL/GenBank/DBJ databases  
DR EMBL: AB062912, BAB60807.1, -  
DR InterPro: IPR004294; RPE65.  
DR Pfam: PF03055; RPE65, 1.  
KW Dioxigenase.  
SQ SEQUENCE 566 AA: 63637 MW: 184367815247892 37674:

Query Match 4.8%, Score 25, DB 11, Length 566,  
Best Local Similarity 100.0%, Pred. No. 2 le-18;  
Matches 25, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 82 NIEANRIVVSEFGTMAYDPCKNIF 106  
|||||  
DB 82 NIEANRIVVSEFGTMAYDPCKNIF 106

## RESULT 7

Q90WH4  
ID Q90WH4 PRELIMINARY: PRT: 516 AA  
AC Q90WH4  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Putative b,b'-carotene-15,15'-dioxigenase.  
GN B-DIOX.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER.  
RX MEDLINE=21216714; PubMed=11278918;  
FA Kiefer C., Hessel S., Lampert M., Vogt K., Lederer M.O.,  
RA Breithaupt D.E., von Lintig J.;  
PT "Identification and characterization of a mammalian enzyme catalyzing  
RT the asymmetric oxidative cleavage of provitamin A".  
RL J. Biol. Chem. 276:14110-14116(2001).  
DR EMBL: AL290590; CAC47566.1, -  
DR InterPro: IPR004294; RPE65.  
DR Pfam: PF03055; RPE65, 1.  
KW Dioxigenase.  
SQ SEQUENCE 516 AA: 58584 MW: 184367815247892 37674:

Query Match 4.8%, Score 13, DB 13, Length 516;  
Best Local Similarity 100.0%, Pred. No. 4.2e-05;  
Matches 13, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 94 GTMAYDPCKNIF 106  
|||||  
DB 94 GTMAYDPCKNIF 106

## RESULT 8

Q9NKG9  
ID Q9NKG9 PRELIMINARY: PRT: 620 AA  
AC Q9NKG9  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE DRPE65.  
GN RPE65 OR RPE65 OR Q9347.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Aphidiroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-P.  
RA Sahara H., Suzuki E.;  
PT "Drosophila cDNA similar to RPE65".  
PL Submitted (APR 2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB041507; BAA94508.1, -  
DR FlyBase: FBgn038171, RPE65.  
DR InterPro: IPR004294; RPE65.  
DR Pfam: PF03055; RPE65, 1.  
KW Dioxigenase.  
SQ SEQUENCE 620 AA: 63647 MW: 184367815247892 37674:

```

Query Match      2.18; Score 11; DB 5; Length 620;
Best Local Similarity 100.0%; Pred. No. 0.0083;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

DR EMBL: AJ276682; CAB93141.1; -
DR Flybase: FBgn0038171; RPE65.
DR InterPro: IPI004294; RPE65.
DR Pfam: PF03055; RPE65.1.
KW Dioxxygenase.
SQ SEQUENCE 520 AA; 69441 MW; 919722EH02C3103 CRC64;

Query Match      2.18; Score 11; DB 5; Length 620;
Best Local Similarity 100.0%; Pred. No. 0.0083;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 453 PSEPIFVPSPD 463
DB 550 PSEPIFVPSPD 560
|||||
RESULT 9
Q9VFS2 PRELIMINARY; PRT; 620 AA.
AC Q9VFS2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG9347 protein (15-15' beta carotene dioxxygenase).
GN RPE65 OR BETA-DIOX OR CG9347
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Insecta;
OC Eukaryota, Neoptera, Euphydrygidae, Lepidoptera, Brachyoptera, Muscomorpha;
OC Ephydroidea, Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132.
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins P.A., Galie P.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin G.,
RA Ballou K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Hurtis K.C., Busam D.A., Butler H., Cadieu F., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport I.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Loudon K., Doup L.E., Downes M., Dudan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisok A., Gong F., Gorelli J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jabali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp P., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy R., Murphy L., Muzey D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein-Dave B., Pauley J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard T., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders P.P.C., Scheeler F., Shen H.,
RA Shue B.C., Staden-Klamis L., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Staden-Klamis A., Stapleton M., Sprung P., Sun F.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z. Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Wu D., Wu D., Wu D., Wu D., Wu D., Wu D., Wu D.,
RA Ye J., Yeh R.-F., Zaveri I.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhang P.N., Zhang W., Zhao X., Zhu S., Zhu X., Smith H.C.,
RA Gibbs K.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000)
RP SEQUENCE FROM N.A.
RX MEDLINE=20224792; PubMed=10766819;
RA von Lintig J., Vogt K.;
RT "Filling the gap in vitamin A research."
RL J. Biol. Chem. 275:11915-11920(2000)
EMBL: AE003701; AAF54978.1; -

```

```

R1 J Biol. Chem. 275:14110-14116(2001).
DR EMBL: AJ290391; CAC37567.1;
DR InterPro: IPR004294; RPE65;
DR Pfam: PF03055; RPE65; 1.
KW Dioxigenase.
SQ SEQUENCE 549 AA; 52426 MW; 6B4FCAD1BF87DBR7 CRC64;

Query Match 1.9%; Score 10; DB 13; Length 549;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NPVVVSFRGT 95
DB 111 NPVVVSFRGT 120

RESULT 12
Q9AEG4 PRELIMINARY; PRT; 169 AA.
AC Q9AEG4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Putative AefA protein (Fragment).
GN AefA.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=548;
RN 111 SEQUENCE FROM N.A.;
RP STRAIN-BW16627;
RC Pradel E., Pages J.M.;
RT "The AcrA/AcrB/TolC efflux pump participates in multidrug resistance
RT in Enterobacter aerogenes."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DP EMBL: AJ306389; CAC35722.1;
FT NON_TER 169
SQ SEQUENCE 169 AA; 19014 MW; 33EBC250079FD0E5 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 OTLETLDK 157
DB 67 OTLETLDK 74

RESULT 13
Q9RJA2 PRELIMINARY; PRT; 287 AA.
AC Q9RJA2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative binding protein dependent transport protein.
GN Streptomyces coelicolor.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN 111 SEQUENCE FROM N.A.;
RP STRAIN-A3(2);
RC Oliver K., Harris D.;
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases
[2]
RN 121 SEQUENCE FROM N.A.;
RP STRAIN-A3(2);
RC Thomson N.R., Parkhill J., Barrrell J.G., Rajandream M.A.;
PA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases
[3]
RN 131 "Functional annotation of a full-length mouse cDNA collection."

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-67000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Gullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996);
RN 141 SEQUENCE FROM N.A.;
RP STRAIN-A3(2) / M145;
RC Bentley S.D., Clater K.F., Cardeno-Liarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark T., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz S., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL132973; CAB61179.1;
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp. 1.
SQ SEQUENCE 287 AA; 31533 MW; 77AB4EDEC0FFA6CF CRC64;

Query Match 1.5%; Score 8; DB 16; Length 287;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 LGRATVNV 505
DB 14 LGRATVNV 21

RESULT 14
Q9CTP3 PRELIMINARY; PRT; 293 AA.
AC Q9CTP3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE A930029L0681k protein (Fragment).
GN A930029L0681K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 111 SEQUENCE FROM N.A.;
RP STRAIN-C57BL/6J; TISSUE=RETINA;
RX MEDLINE-z1085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gajewski J., Ezoe H., Kasukawa I., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland P., Gissi C., King K., Kochiwa H.,
RA Kuchii P., Lewis S., Matsuo Y., Nakado I., Resole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gastinech S., Hill D., Hofmann M., Hume E.A., Kamiya M., Lee N.H.,
PA Lyons P., Marchionni L., Mashima J., Mazzarelli J., McBeris P.,
RA Nordone P., Ping R., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
PA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
PA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."

```

RL Nature 409:685-690(2001).  
 DR EMBL: AK020906; BAB32248.1; -.  
 DR MCD: MGI:192506R; Aqinn2q106Rik.  
 DR InterPro: IPR004294; RPE65.  
 DR Pfam: PF03055; RPE65; 1.  
 FT NCON\_TER 293 293  
 SQ SEQUENCE 293 AA; 33281 MW; F4B867CDEF4A3E7D CRC64;

Query Match 1.5%; Score 8; DR 11; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 8.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PPCKNIF 106  
 DB 109 PPCKNIF 116

## RESULT 15

Q95RP9 PRELIMINARY; PRT; 333 AA.  
 AC G95RP9;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE LD16758p.  
 GN OCT OF PCTNA-GH04245 OF Cct10392  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Aqbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George P.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Gelniker S.;  
 RC Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY061225; AA128773.1; -.  
 DR Flybase: FBgn0040295; oct.  
 SQ SEQUENCE 333 AA; 37253 MW; 9c75223e506d9bf9 CRC64;

Query Match 1.5%; Score 8; DR 5; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 KIDPQTL 153  
 DB 146 KIDPQTL 153

Search completed: July 15, 2003, 09:42:43  
 Job time : 86 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 15, 2003, 21:28:47, Search time: 787 seconds  
(without alignments)  
11502.837 Million coll updates/sec  
Title: US-10-053-192-2  
Perfect score: 3111  
Sequence: 1 cggatccactagtaacggcc aaaaaaaaaaaaaaaaaaaaaa 3111

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0  
Searched: 2054640 seqs, 14551402878 residues  
Word size: 0  
Total number of hits satisfying chosen parameters: 4109,80  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database		GenBank	
1:	gb_fa *	1:	gb_fa *
2:	gb_htg *	2:	gb_htg *
3:	gb_in *	3:	gb_in *
4:	gb_om *	4:	gb_om *
5:	gb_ov *	5:	gb_ov *
6:	gb_pat *	6:	gb_pat *
7:	gb_ph *	7:	gb_ph *
8:	gb_pi *	8:	gb_pi *
9:	gb_pr *	9:	gb_pr *
10:	gb_ro *	10:	gb_ro *
11:	gb_sts *	11:	gb_sts *
12:	gb_sy *	12:	gb_sy *
13:	gb_un *	13:	gb_un *
14:	gb_vi *	14:	gb_vi *
15:	em_b4 *	15:	em_b4 *
16:	em_fun *	16:	em_fun *
17:	em_hum *	17:	em_hum *
18:	em_in *	18:	em_in *
19:	em_mu *	19:	em_mu *
20:	em_mn *	20:	em_mn *
21:	em_or *	21:	em_or *
22:	em_ov *	22:	em_ov *
23:	em_pat *	23:	em_pat *
24:	em_ph *	24:	em_ph *
25:	em_pl *	25:	em_pl *
26:	em_ro *	26:	em_ro *
27:	em_sts *	27:	em_sts *
28:	em_un *	28:	em_un *
29:	em_vi *	29:	em_vi *
30:	em_htg_hum *	30:	em_htg_hum *
31:	em_htg_inv *	31:	em_htg_inv *
32:	em_htg_dti *	32:	em_htg_dti *
33:	em_htg_mus *	33:	em_htg_mus *
34:	em_htg_pla *	34:	em_htg_pla *
35:	em_htg_rod *	35:	em_htg_rod *
36:	em_htg_wam *	36:	em_htg_wam *
37:	em_htg_vrt *	37:	em_htg_vrt *
38:	em_sy *	38:	em_sy *
39:	em_htg_hum *	39:	em_htg_hum *
40:	em_htg_mus *	40:	em_htg_mus *
41:	em_htg_other *	41:	em_htg_other *

Pred No is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3111	100.0	3111	5	GA271386
2	3111	100.0	3111	5	AX033822 Sequence
3	3111	100.0	3111	5	AX033822 Sequence
4	36	1.2	219	6	AX112509 Sequence
5	36	1.2	219	6	AX112509 Sequence
6	36	1.2	219	6	AX157579 Sequence
7	36	1.2	219	6	AX201985 Sequence
8	36	1.2	1524	10	RNEGP314H
9	36	1.2	1730	8	AF073473
10	36	1.2	1982	6	AF079853
11	36	1.2	1984	8	AF015523
12	35	1.1	1076	6	AF010163
13	35	1.1	1302	6	AF023084
14	34	1.1	3130	9	HS029770
15	33	1.1	870	6	AX157785
16	30	1.0	278	6	AX137601
17	30	1.0	355	6	AX175602
18	30	1.0	3208	9	HS245567
19	29	0.9	185224	3	AF097817
20	29	0.9	101	6	AX284907
21	29	0.9	103	6	AX197644
22	29	0.9	114	6	AX361301
23	29	0.9	118	6	AX051505
24	29	0.9	118	6	AX072748
25	29	0.9	118	6	AX074190
26	29	0.9	122	6	AX442214
27	29	0.9	143	6	AX089604
28	29	0.9	147	6	AX167247
29	29	0.9	151	6	AX172437
30	29	0.9	170	6	AX167246
31	29	0.9	174	6	AX261864
32	29	0.9	186	6	AX198908
33	29	0.9	195	6	AX198902
34	29	0.9	198	6	AX198918
35	29	0.9	201	6	AX198909
36	29	0.9	203	9	HS036623
37	29	0.9	207	6	AX198656
38	29	0.9	210	6	AX263556
39	29	0.9	211	6	AX198641
40	29	0.9	212	6	AX198571
41	29	0.9	217	6	AX198741
42	29	0.9	220	6	AX198825
43	29	0.9	222	6	AX375606
44	29	0.9	222	6	AX455841
45	29	0.9	223	6	AX198317
			226	6	AX198212

ALIGNMENTS

RESULT 1  
GG271386  
LOCUS  
DEFINITION  
Gallus gallus mRNA for beta-carotene 15,15'-dioxygenase (bCDO gene).  
3111 bp mRNA linear VRT 23-MAY-2001  
ACCESSION  
AJ271386  
VERSION  
bCDO gene; beta-carotene 15,15'-dioxygenase.  
KEYWORDS  
chicken.  
SOURCE  
Gallus gallus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE  
1 (bases 1 to 3111)  
AUTHORS  
Wyss, A., Wirth, G., Weegen, W., Progg, P., Wyss, M., Friedlein, A.,



[illegible]

Db 2401 ATGTAATACCAAAATTCTGATAAATCAGTTTCCTTTCAGAAATATTCAGACAACATTTT 2460  
 QY 2461 GTCTAAATCAATCAATATATCTTTTAAATGAATCAATATATCTTTTATAGTATTTG 2520  
 Db 2461 GTCTAAATCAATCAATATATCTTTTAAATGAATCAATATATCTTTTATAGTATTTG 2520  
 QY 2521 TTAATGTCTTACTATATCTATCATATGTAANTGAGAGCAATGTATCTTACGAGAAGCTCA 2580  
 Db 2521 TTAATGTCTTACTATATCTATCATATGTAANTGAGAGCAATGTATCTTACGAGAAGCTCA 2580  
 QY 2581 GATATACATTCACAACATTTCTGTAGGTGAAAATGCCATTTTACTGATCAAAATCTCAATGGTT 2640  
 Db 2581 GATATACATTCACAACATTTCTGTAGGTGAAAATGCCATTTTACTGATCAAAATCTCAATGGTT 2640  
 QY 2641 AATCAAGAGCAAAAAATCGGGTAATCCATATCCCAACTATGTTAGTGTTTACCTGGCTGTGT 2700  
 Db 2641 AATCAAGAGCAAAAAATCGGGTAATCCATATCCCAACTATGTTAGTGTTTACCTGGCTGTGT 2700  
 QY 2761 ATGTCGACACACACCTGTTTGGSHATCCATCTTCACATACATGTCGTCACAGGTTCTCG 2760  
 Db 2761 ATGTCGACACACACCTGTTTGGSHATCCATCTTCACATACATGTCGTCACAGGTTCTCG 2760  
 QY 2761 CTATCTACTGTATTTTCAAGAGAGAGCAAAAAATCAATCACTCTCACTTATATAGAGGACC 2820  
 Db 2761 CTATCTACTGTATTTTCAAGAGAGAGCAAAAAATCAATCACTCTCACTTATATAGAGGACC 2820  
 QY 2821 CATCAATATAGAACATATGGGGGCTTTGGTGACTTGTTCATATTAATCAATCATGTTTCTAG 2880  
 Db 2821 CATCAATATAGAACATATGGGGGCTTTGGTGACTTGTTCATATTAATCAATCATGTTTCTAG 2880  
 QY 2881 CAGAGCAGCAATAGAACCAACACATGCACAGTTCTTGGCTGGCTGTATATGATTCGCTT 2940  
 Db 2881 CAGAGCAGCAATAGAACCAACACATGCACAGTTCTTGGCTGGCTGTATATGATTCGCTT 2940  
 QY 2941 TGCTGTCTTTATGTTTTGATGTAAGAGAAATACATGCTTAATCTTAATGTTTAAAAAT 3000  
 Db 2941 TGCTGTCTTTATGTTTTGATGTAAGAGAAATACATGCTTAATCTTAATGTTTAAAAAT 3000  
 QY 3001 CATCTGGGCTCAGATCTAGAGCTTAAGSTTAAGCAGCTGGGCTTTTCCAATGTTTATATGTT 3060  
 Db 3001 CATCTGGGCTCAGATCTAGAGCTTAAGSTTAAGCAGCTGGGCTTTTCCAATGTTTATATGTT 3060  
 QY 3061 CCATATAATGGAATAAACAACACATCCATTAATAAAAAAAAAAAAAAAAAAAAA 3111  
 Db 3061 CCATATAATGGAATAAACAACATCCATTAATAAAAAAAAAAAAAAAAAAAAA 3111

```

RESULT 2
AX033822      AX033822      3111 bp      DNA      110bp      PA: 21-SEP-2009
LOCUS        Sequence 2 from Patent EP1031627
DEFINITION   AX033822
ACCESSION   AX033822
VERSION     AX033822.1 GI:19280427
KEYWORDS
SOURCE       chicken.
ORGANISM     Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria, Aves; Neognathae, Galliformes, Phasianidae;
Phasianinae; Gallus.
REFERENCE    1 (bases 1 to 3111)
AUTHORS     Bachmann,H., Wyss,A., Wyss,M., Woggon,W.D., Friedlein,A.M.,
            Wirtz,G.M. and Bruegger,R.
TITLE        Beta-carotene 15,15'-dioxygenase
JOURNAL     Proc Natl Acad Sci USA 96(1999);
HOFFMANN LA ROCHE (CH)
FEATURES     Location/Qualifiers
             source          1..3111
                        /organism="Gallus gallus"
                        /db_xref="taxon:9031"
BASE COUNT   1004 a 1004 c 1022 g 864 t
ORIGIN
Query Match      100.0%; Score 3111; DB 6; Length 3111;
  
```





[illegible]

1621	Db	GAATTGGGGCGAGCCACAGTTAACTGACAAATGCATCTGCAGCTGATGCATGTTTATA	1680
1681	Qy	CCACAGAAATGATTGGGGGCTGACAGCAATAAAGTCTATTGATGGAAACACACAAAT	1740
1681	Db	CCACATAATGATTTGGGGGCTGACATAGCAAAACAAACGGCTATGATCGGACTACACAACT	1740
1741	Qy	GAGACAACCTTCTACGACCAATGAGUUAATATGCTTTTACGATTTCAGAAATTAAT	1800
1741	Db	GAGACAACCTTCTACTCAAACTACGTTTAATATGCTTTTACGATTTCAGAAATTAAT	1800
1801	Qy	AAGGACACAAAATGCACTATCTATAAATCTCTTAATAATAACATATAATCTTTTAAACAC	1860
1801	Db	AAGGACACAAAATGCACTATGATAAATCTCTTAATAATAACATATAATCTTTTAAACAC	1860
1861	Qy	AGCGATGASCTTTTACTACAGGTAACGATATGCAAACTCTGGCATATAACTATTTC	1920
1861	Db	AGGCAAGAGTTTACACAGGTAACCAATATGCAAACTCTGGCATATAAATATTTC	1920
1921	Qy	AGAAGAACGATGAGTGTGTTTTAGAAAGTGCTTAATCTTGATACATACAGGGGCGATAGGAAT	1980
1921	Db	AGAAGAACGATGAGTGTGTTTTAGAAAGTGCTTAATCTTGATACATACAGGGGCGATAGGAAT	1980
1981	Qy	GGAGAGAAAGGTAAACGGGAATATTTAAACAAATATAGATTTCTGAGCAATGAAGTGAC	2040
1981	Db	GGAGAGAAAGGTAAACGGGAATATTTAAACAAATATAGATTTCTGAGCAATGAAGTGAC	2040
2041	Qy	TATTTATGTTGATGATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2100
2041	Db	TATTTATGTTGATGATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2100
2101	Qy	GATCGTTCAAGATTCGAGCTTCGATGCAAGTTCCTCCAGGCGACAGAAACGTCATTTA	2160
2101	Db	GATCGTTTAAAGATTCGAGCTTCGATGCAAGTTCCTCCAGGCGACAGAAACGTCATTTA	2160
2161	Qy	AACCATGCTATGGTATTAATCAATCAATGCAATTTCTCTGGGCGCATTAACATATA	2220
2161	Db	AACCATGCTATGGTATTAATCAATCAATGCAATTTCTCTGGGCGCATTAACATATA	2220
2221	Qy	ACCAAGCTTANGATTACATTCAGTGGGCGACAACTTCTCAATTTACAAACCAACAA	2280
2221	Db	ACCAAGCTTANGATTACATTCAGTGGGCGACAACTTCTCAATTTACAAACCAACAA	2280
2281	Qy	AGCAACAGCACTTGCTTGGCTAAACCGCATGGTGCTATTTTCTCTTTTATGATGACA	2340
2281	Db	AGCAACAGCACTTGCTTGGCTAAACCGCATGGTGCTATTTTCTCTTTTATGATGACA	2340
2341	Qy	AAACCAAGTACATATGTTTTATCTAGCATTCATTTATCTTACGTCATTCATTCATTA	2400
2341	Db	AAACCAAGTACATATGTTTTATCTAGCATTCATTTATCTTACGTCATTCATTCATTA	2400
2401	Qy	ATGTTATAAGCAATTTGATTTAAATCAGTTTTCTCTTGAGAAATATCTGACATAAATTT	2460
2401	Db	ATGTTATAAGCAATTTGATTTAAATCAGTTTTCTCTTGAGAAATATCTGACATAAATTT	2460
2461	Qy	GTGTAATGAGTACATATGTTGTTCTTAAGATGAACAGCAATCTATCTTTTATGATTTG	2520
2461	Db	GTGTAATGAGTACATATGTTGTTCTTAAGATGAACAGCAATCTATCTTTTATGATTTG	2520
2521	Qy	TTAATTTGTTTACTAATCTATGTCATATAAATGAGCAATGATTTTACGACAAATCA	2580
2521	Db	TTAATTTGTTTACTAATCTATGTCATATAAATGAGCAATGATTTTACGACAAATCA	2580
2581	Qy	GATATACATCAACAAATTTCTGAGTGGAATATGCAATTTACTGATGAAGATGCAATGTT	2640
2581	Db	GATATACATCAACAAATTTCTGAGTGGAATATGCAATTTACTGATGAAGATGCAATGTT	2640
2641	Qy	AATGAGGAGAAAGTGGGATATCATGCAATGCAATATGTTACGCTTCAGCTGGCTGCT	2700
2641	Db	AATGAGGAGAAAGTGGGATATCATGCAATGCAATATGTTACGCTTCAGCTGGCTGCT	2700
2701	Qy	ATGTGACACAGCTGTTTGGGTATGCTCTACCTTTTCACATATGCTGCTGATGCTTCTG	2760



```

/codon_line="regressor"
/tissue_type="colon carcinoma"
/codon_line="oligo-dt-primed cDNA in pcDNA3"
123..1070
/codon_start=1
/product="Esp-314 protein homologue"
/protein_id="CAA04498.1"
/db_xref="GI:2764585"
/db_xref="SPIRHH0:055154"
/translation="MAPKKALPGLIIAVVTIIAAAGKQVWNNYKILSMYFNFN;
ECCTSYGNTIVISKLASKLVMKAEHTSGSRMRKPEGAIONDGLYDPECDEQ
SLFKACQNTATCWCNTAGVPTDKDEITCSEPVFTYIIELKHKEAPQYNEF
GLTALQDTPASRYMLNPKFIKSIYNNVITIDLMONSOKTODDVIDADVAYFEK
DVKGSLSFHSKSMOLRVNGELLDIDPGGILLIYVDEKAPESMGLTAGIIVVV
VLAVIAGIVVLVISTHRSKRYKAFIKEMGEIHELNA"
BASE COUNT 417 a 339 c 411 g 357 t
ORIGIN
Query Match 1.2%; Score 36; DB 10; Length 1524;
Best Local Similarity 100.0%; Pred. No. 7.3e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGGCGGCGAGTGGTGGAAT 36
|||||
DB 18 CGGATCATTAGTAAAGGGCGGCGAGTGGTGGAAT 53
|||||

RESULT 8
AF073473
LOCUS Solanum tuberosum 1730 bp mRNA linear PIN 18-JUL-1998
DEFINITION Solanum tuberosum phospholipidyl transferase kinase precursor, mRNA, nuclear
VERSION AF073473
KEYWORDS Solanum tuberosum; Solanaceae; Solanum
SOURCE Thorbjornsen, T.
ORGANISM Solanum tuberosum
REFERENCE 1 (bases 1 to 1730)
AUTHORS Thorbjornsen, T.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1998) Department of Plant Biology, Royal
Veterinary and Agricultural University, Thorvaldsensvej 40,
Frederiksberg C 1871, Denmark
FEATURES
source 1..1730
/organism="Solanum tuberosum"
/db_xref="taxon:4113"
/tissue_type="leaf"
209..1657
/codon_start=1
/product="phosphoglycerate kinase precursor"
/db_xref="GI:3328122"
/translation="MASATASHFTFCGIPITSSSTTNRSTTHSSARFLKTPRLRQF
AGAVAPLEINVAIKRLSKASSNPRAVSMARKSVGLSSDLKGVVVRADLN
VPLDQANTDDTRIPAAIPTIKHIANCAKVIISSHLGKFGVTPKYSIAPLPLS
ELLGIQVADKDCIGPEVEKLVDSLPKGVLLLENVREYKEEKNEPEAKKLASLAD
LFVNDATQAHRAHASTEGVTKFKPSVAGFLKELDYLVAVSTPKPPPAATVGS
KVSKIGVTESLLEKCDIILGGMIFTFYKAGLSVGSLSVEEDKLELATSLLEKAK
AKGVSLIPSDVVIAUKPAPDANSKIVPASAIPDQWMLDIPGDSIKTFNDALDTKT
VINHGQVFEPEKATGTEATKRLADLSRGVITIIIGGDSVAAEKVGASVMSH
ISTGACGELLEKGVLPGLVIALDADAPAV"
BASE COUNT 449 a 400 c 406 g 475 t
ORIGIN
Query Match 1.2%; Score 36; DB 8; Length 1730;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 36; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCATTAGTAAAGGGCGGCGAGTGGTGGAAT 36
|||||

```

```

DB 29 CGGATCATTAGTAAAGGGCGGCGAGTGGTGGAAT 64
|||||

RESULT 9
AF079853
LOCUS 1982 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5965792.
VERSION AF079853
KEYWORDS 1 GI:10006594
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1982)
AUTHORS Schroeder, J.I., Antosiewicz, D.M., Schachtman, D.P. and Clemens, S.
TITLE Nucleic acids encoding metal uptake transporters and their uses
JOURNAL Patent: US 5965792-A 1 12-OCT-1999;
FEATURES
source 1..1982
Location/Qualifiers
source 354 a 611 c 569 g 448 t
/organism="unknown"
Query Match 1.2%; Score 36; DB 6; Length 1982;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCATTAGTAAAGGGCGGCGAGTGGTGGAAT 36
|||||
DB 18 CGGATCATTAGTAAAGGGCGGCGAGTGGTGGAAT 53
|||||

RESULT 10
AF015523
LOCUS 1984 bp mRNA linear PIN 29-SEP-1999
DEFINITION Triticum aestivum low-affinity cation transporter (LCT1) mRNA,
complete cds.
VERSION AF015523
KEYWORDS 1 GI:2460043
SOURCE Triticum aestivum.
ORGANISM Triticum aestivum
REFERENCE 1 (bases 1 to 1984)
AUTHORS Schachtman, D.P., Kumar, R., Schroeder, J.I. and Marsh, E.L.
TITLE Molecular and functional characterization of a novel low affinity
cation transporter (LCT1) in higher plants
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (20), 11079-11084 (1997)
MEDLINE 98021496
PUBMED 9380762
REFERENCE 2 (bases 1 to 1984)
AUTHORS Schachtman, D.P., Kumar, R., Schroeder, J.I. and Marsh, E.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1997) Botany, University of Adelaide, Adelaide,
SA 5005, Australia
FEATURES
source 1..1984
Location/Qualifiers
source 1..1984
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/tissue_type="root"
/feature="cloned from cDNA library by yeast complementation"
1..1984
/gene="LCT1"
137..1861
/gene="LCT1"
/codon_start=1
/product="low-affinity cation transporter"
/protein_id="AAB71851.1"
/db_xref="ref:2460043"

```





## RESULT 14

AR157785  
 LOCUS AR157785 870 bp DNA linear PA: 17-OCT-2001  
 DEFINITION Sequence 1 from patent US 6245898.  
 ACCESSION AR157785  
 VERSION AR157785.1 GI:16218795  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 870)  
 Testa, J.E., Quigley, J.P. and Seandel, M.  
 Monoclonal antibodies that recognize antigens associated with tumor  
 metastasis  
 JOURNAL Patent: US 6245898-A 1 12-JUN 2001;  
 FEATURES  
 location/Qualifiers  
 source  
 1..870  
 /organism="unknown"  
 BASE COUNT 168 a 253 c 262 g 187 t  
 ORIGIN

Query Match 1.1%; Score 33; DB 6; Length 870;

Best Local Similarity 100.0%; Pred. No. 3.5e-06;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATCCACTAGTAACGGCGGCAGTGTGGTGGAAAT 36

|||||

Db 1 ATCCACTAGTAACGGCGGCAGTGTGGTGGAAAT 33

## RESULT 15

AX175601/C  
 LOCUS AX175601 275 bp DNA linear PAT 03-JUL-2001  
 DEFINITION Sequence 57 from Patent WO0144512.  
 ACCESSION AX175601  
 VERSION AX175601 1 GI:14598921  
 KEYWORDS  
 SOURCE Norway Rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 275)  
 Gould-Rothberg, B.E.  
 Method of identifying ligands for the peroxisome proliferator  
 activated receptor gamma using differential gene expression  
 Patent: WO 0144512-A 57 21-JUN-2001;  
 JOURNAL Curagen Corporation (US)  
 FEATURES  
 location/Qualifiers  
 source  
 1..275  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"

BASE COUNT 62 a 74 c 79 g 57 t 3 others

## ORIGIN

Query Match 1.0%; Score 30; DB 6; Length 275;

Best Local Similarity 100.0%; Pred. No. 0.00018;

Matches 30; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAACGGCGGCAGTGTGG 30

|||||

Db 169 CGGATCCACTAGTAACGGCGGCAGTGTGG 140

Search completed: July 16, 2003, 02:26:10

Job time : 7875 secs

GenCore version 5.1.6  
Copyright (c) 1994 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 23:55:37 ; Search time 4167 seconds  
(without alignments)  
12091.231 Million cell updates/sw

Title: US-10-053-192-2

Perfect score: 3111

Sequence: 1 cggatccactagtaacggcc... aaaaaaaaaaaaaaaaaaaaaa x111

Scoring table: OLIGO-NOC

Gapop 60 0 , Gapext 60.0

Searched: 16154066 seqs, 809774376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32309132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : FST\*\*

1:	em_estba:**
2:	em_esthm:**
3:	em_estm:**
4:	em_estmu:**
5:	em_estov:**
6:	em_estpl:**
7:	em_estro:**
8:	em_hic:**
9:	gb_est1:**
10:	gb_est2:**
11:	gb_hic:**
12:	gb_est3:**
13:	gb_est4:**
14:	gb_est5:**
15:	em_esthm:**
16:	em_estom:**
17:	gb_ysc:**
18:	em_ysc_hum:**
19:	em_ysc_inv:**
20:	em_ysc_pin:**
21:	em_ysc_vrt:**
22:	em_ysc_fun:**
23:	em_ysc_mam:**
24:	em_ysc_mus:**
25:	em_ysc_other:**
26:	em_ysc_pro:**
27:	em_ysc_rdi:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586	18.8	661	13	B1389779 papic.pk0
2	36	1.2	73	9	AL449724 AL449724
3	36	1.2	197	9	AJ133823 AJ133823
4	36	1.2	259	14	BM887893 BM887893
5	36	1.2	386	9	AA514191 AA514191
6	36	1.2	418	14	W91597 W91597

7	36	1.2	434	17	AA514191
8	36	1.2	451	9	AL449724
9	36	1.2	452	10	BE414749
10	36	1.2	458	13	B1203079
11	36	1.2	496	17	A2251292
12	36	1.2	503	9	AA514191
13	36	1.2	555	14	W91597
14	36	1.2	612	14	BM887893
15	36	1.2	633	14	BM887893
16	36	1.2	644	10	AW609004
17	33	1.1	464	10	BE414750
18	33	1.1	475	10	BE414889
19	32	1.0	465	10	BE414893
20	32	1.0	472	14	BM887893
21	31	1.0	417	10	BE414879
22	31	1.0	467	10	BE414892
23	30	1.0	285	14	W91597
24	30	1.0	768	14	HQ747840
25	29	0.9	102	10	AW608988
26	29	0.9	103	12	BE927286
27	29	0.9	106	10	AW609004
28	29	0.9	108	10	AW608994
29	29	0.9	109	13	AW937014
30	29	0.9	110	10	AW608994
31	29	0.9	113	12	BE286784
32	29	0.9	111	10	AW508997
33	29	0.9	111	10	AW879277
34	29	0.9	111	12	BE286784
35	29	0.9	113	10	AW608996
36	29	0.9	114	10	AW577820
37	29	0.9	115	10	AW947006
38	29	0.9	117	10	AW608983
39	29	0.9	121	10	AW608988
40	29	0.9	121	10	AW609018
41	29	0.9	123	10	AW609004
42	29	0.9	124	10	AW609003
43	29	0.9	124	12	BE927285
44	29	0.9	125	10	AW936837
45	29	0.9	125	10	AW947006

#### ALIGNMENTS

RESULT 1  
B1389779  
Library: papic.pk002.ali2 Primary chicken Ptiliary-Hypothalamus/pineal  
Library: Gallus gallus cDNA clone papic.pk002.ali3 5' similar to  
emb|CA60985.1 (ATG7196) beta-carotene 15,15'-dioxygenase [Gallus  
gallus], mRNA sequence.  
B1389779  
B1389779.1 GI:15083061  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Gallus gallus  
Chicken.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
Porter,T.E. and Coqburn,I.A.  
ESTs from Primary Chicken Ptiliary/Hypothalamus/Pineal cDNA  
Library USDA/IFAPS Animal Genome Project  
Unpublished (2001)  
Contact: Larry A. Coqburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1345  
Fax: 302-831-2822  
Email: coqburn@udel.edu, www.chickest.udel.edu.  
FEAT:FEFS  
Location/Qualifiers  
1..661  
/organism="Gallus gallus"





Best Local Similarity 100.0%; Pred. No. 0;  
Matches 36; Conservative 0; Mismatches 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 36  
|||||  
DB 423 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 288

RESULT 7  
LOCUS AZ254240 434 bp DNA linear GSS 21 JUN 2000  
DEFINITION mRM2P RFLP sequences of mungbean, Vigna radiata  
ACCESSION AZ254240  
VERSION AZ254240.1 GI:8602504  
KEYWORDS GSS.  
SOURCE Vigna radiata  
ORGANISM Vigna radiata  
REFERENCE 1 (bases 1 to 434)  
AUTHORS Denny, E., Damesh, D., Madec, J., Cooper, A., Larson, K.,  
Menancio-Hautea, D., Kumar, L. and Young, N. D.  
TITLE RFLP sequences of mungbean, Vigna radiata  
JOURNAL Unpublished (2000)  
COMMENT Contact: Young Nevin D.  
Department of Plant Pathology  
University of Minnesota  
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA  
Tel: 612 625 2225  
Fax: 612 625 9728  
Email: nevin@tc.umn.edu

Sequence of mapped RFLP marker mRM2 on linkage group I of  
mungbean; linkage group II of cowpea. For more information, see  
Beangenes at:  
<http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=beangenes&cl=ass=locus>. Please see authorities for mapping/naming: Fatokun,  
C.A., Damesh, D., Young, N.D. (1993) RFLP linkage map for cowpea  
(Vigna unguiculata (L.) Walp.) in: Genetic Maps 1992, S. J.  
O'Brien, ed. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp.  
6.256-6.258.  
Young, N.D. (1993) RFLP linkage map for mungbean (Vigna radiata (L.)  
Wilczek) in: Genetic Maps 1992, S. J. O'Brien, ed. Cold Spring  
Harbor Press, Cold Spring Harbor, NY, pp. 6.259-6.260  
Insert length: 1300 Std Error: 0.00  
Seq primer: M138  
Class: RFLP probe.

FEATURES  
source  
1..434  
/organism="Vigna radiata"  
/cultivar="Textsprout"  
/db\_xref="taxon:157791"  
/clone\_lib="RFLP sequences of mungbean, Vigna radiata"  
/tissue\_type="Hypocotyl and roots"  
/dev\_stage="Sprouts"  
/note="Vector: pUC 18; Site 1. Pst I. DNA was digested  
with Pst I, size separated by sucrose gradient  
centrifugation and the fraction between 500-3000 base  
pairs ligated into the vector using standard method."

BASE COUNT 103 a 115 c 101 g 101 t 14 others  
ORIGIN

Query Match 1.2%, Score 36; DB 17; Length 434;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 36  
|||||  
DB 18 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 53

RESULT 8  
LOCUS AL449722 451 bp mRNA linear EST 15 NOV 2000  
DEFINITION AL449722 Homo sapiens fetal brain (Stavrides GS) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION AL449722  
VERSION AL449722.1 GI:11191356  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Chordata; Cladocera; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 451)  
AUTHORS Stavrides, G.S., Huckle, E.J. and Deloukas, P.  
TITLE Unpublished. Stavrides, G.S., Huckle, E.J. and Deloukas, P.  
JOURNAL Unpublished (2000)  
COMMENT Contact: Stavrides GS  
The Sanger Centre  
Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: huckle@anger.sanger.ac.uk  
Sanger Centre name: sc001019.47.2fp.

FEATURES  
source  
1..451  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="20"  
/clone\_lib="Homo sapiens fetal brain (Stavrides GS)"  
/tissue\_type="Brain"  
/dev\_stage="fetal"  
/note="cDNA fragment isolated using a cDNA end rescue  
technique"

BASE COUNT 63 a 147 c 135 g 108 t  
ORIGIN

Query Match 1.2%, Score 46; DB 9; Length 451;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 36  
|||||  
DB 12 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGAAT 47

RESULT 9  
LOCUS BE414749 452 bp mRNA linear EST 24 JUL 2000  
DEFINITION BE414749 MML002 A06P90430 ITCF MWL wheat Root Library Triticum aestivum  
cDNA clone MWL002.A06, mRNA sequence.  
ACCESSION BE414749  
VERSION BE414749.1 GI:9412651  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;  
Triticeae; Triticum.  
1 (bases 1 to 452)  
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Hake, J., Close, T., Cloutier  
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,  
Herrmann, P.G., Holton, T., Jacquemin, J.M., Jia, J., Joudriet, P.,  
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogilhard, Y.,  
Perich, N., Quiset, C., Schuch, W., Solvar, G., Sharif, M.,  
Sorrells, M., Warburton, M. and Wenzel, G.

International Triticeae EST Cooperative (ITEC): Production of  
Expressed Sequence Tags for Species of the Triticeae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT  
Apo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7257558/59  
Email: mwarburton@cnet.com  
International Triticeae EST Cooperative (ITEC)

```

http://wheat.pw.usda.gov/genome/
FEATURES             source
  Location/Qualifiers
    1..452
      /organism="Triticum aestivum"
      /cultivar="Atlas"
      /db_xref="taxon:4565"
      /clone="MWL002.A06"
      /clone_lib="ITEC MWL Wheat Root Library"
      /tissue_type="root"
      /dev_stage="8 day old"
      /note="vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average
      insert size."
BASE COUNT          7+ a 120 c 154 g 6+ t 28 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 10; Length 452;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 36
DB 4 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 39
RESULT 10
LOCUS B1203079
DEFINITION MC38R Apple cDNA Library Malus x domestica cDNA clone MC28, mRNA
ACCESSION B1203079
VERSION B1203079.1 GI:14669051
KEYWORDS EST.
SOURCE apple tree.
ORGANISM Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE
AUTHORS 1 (bases 1 to 458)
TITLE Sequences of RFLP mapped cDNAs in Prunus
JOURNAL Unpublished (2001)
COMMENT Contact: Graziano F
Plant Genetics Dpt.
IRTA
C/Cabrils S/N, 08348 Cabrils (Barcelona), Spain
Email: Pere.Arusauberta.es.
FEATURES             source
  Location/Qualifiers
    1..458
      /organism="Malus x domestica"
      /db_xref="taxon:3750"
      /clone="MC28"
      /clone_lib="Apple cDNA Library"
      /note="m27 rootstock"
BASE COUNT          95 a 130 c 123 g 10+ t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 13; Length 458;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 45
DB 19 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 54
RESULT 11
LOCUS A2254292
DEFINITION mgQ117R RFLP sequences of mungbean, Vigna radiata
ACCESSION A2254292
VERSION A2254292.1 GI:8602610
KEYWORDS GSS.
http://wheat.pw.usda.gov/genome/
FEATURES             source
  Location/Qualifiers
    1..452
      /organism="Triticum aestivum"
      /cultivar="Atlas"
      /db_xref="taxon:4565"
      /clone="MWL002.A06"
      /clone_lib="ITEC MWL Wheat Root Library"
      /tissue_type="root"
      /dev_stage="8 day old"
      /note="vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average
      insert size."
BASE COUNT          7+ a 120 c 154 g 6+ t 28 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 10; Length 452;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 36
DB 4 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 39
RESULT 10
LOCUS B1203079
DEFINITION MC38R Apple cDNA Library Malus x domestica cDNA clone MC28, mRNA
ACCESSION B1203079
VERSION B1203079.1 GI:14669051
KEYWORDS EST.
SOURCE apple tree.
ORGANISM Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE
AUTHORS 1 (bases 1 to 458)
TITLE Sequences of RFLP mapped cDNAs in Prunus
JOURNAL Unpublished (2001)
COMMENT Contact: Graziano F
Plant Genetics Dpt.
IRTA
C/Cabrils S/N, 08348 Cabrils (Barcelona), Spain
Email: Pere.Arusauberta.es.
FEATURES             source
  Location/Qualifiers
    1..458
      /organism="Malus x domestica"
      /db_xref="taxon:3750"
      /clone="MC28"
      /clone_lib="Apple cDNA Library"
      /note="m27 rootstock"
BASE COUNT          95 a 130 c 123 g 10+ t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 13; Length 458;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 45
DB 19 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 54
RESULT 11
LOCUS A2254292
DEFINITION mgQ117R RFLP sequences of mungbean, Vigna radiata
ACCESSION A2254292
VERSION A2254292.1 GI:8602610
KEYWORDS GSS.
http://wheat.pw.usda.gov/genome/
FEATURES             source
  Location/Qualifiers
    1..452
      /organism="Triticum aestivum"
      /cultivar="Atlas"
      /db_xref="taxon:4565"
      /clone="MWL002.A06"
      /clone_lib="ITEC MWL Wheat Root Library"
      /tissue_type="root"
      /dev_stage="8 day old"
      /note="vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average
      insert size."
BASE COUNT          7+ a 120 c 154 g 6+ t 28 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 10; Length 452;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 36
DB 4 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 39
RESULT 10
LOCUS B1203079
DEFINITION MC38R Apple cDNA Library Malus x domestica cDNA clone MC28, mRNA
ACCESSION B1203079
VERSION B1203079.1 GI:14669051
KEYWORDS EST.
SOURCE apple tree.
ORGANISM Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE
AUTHORS 1 (bases 1 to 458)
TITLE Sequences of RFLP mapped cDNAs in Prunus
JOURNAL Unpublished (2001)
COMMENT Contact: Graziano F
Plant Genetics Dpt.
IRTA
C/Cabrils S/N, 08348 Cabrils (Barcelona), Spain
Email: Pere.Arusauberta.es.
FEATURES             source
  Location/Qualifiers
    1..458
      /organism="Malus x domestica"
      /db_xref="taxon:3750"
      /clone="MC28"
      /clone_lib="Apple cDNA Library"
      /note="m27 rootstock"
BASE COUNT          95 a 130 c 123 g 10+ t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 13; Length 458;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 45
DB 19 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 54
RESULT 11
LOCUS A2254292
DEFINITION mgQ117R RFLP sequences of mungbean, Vigna radiata
ACCESSION A2254292
VERSION A2254292.1 GI:8602610
KEYWORDS GSS.

```

```

Vigna radiata
Vigna radiata
SOURCE
ORGANISM
  Location/Qualifiers
    1..496
      /organism="Vigna radiata"
      /cultivar="Texsprout"
      /db_xref="taxon:157791"
      /clone_lib="RFLP sequences of mungbean, Vigna radiata"
      /tissue_type="Hypocotyl and roots"
      /dev_stage="Sprouts"
      /note="Vector: pUC18; Site 1: Pst I; DNA was digested
      with Pst I; size separated by sucrose gradient
      centrifugation and the fraction between 500-3000 base
      pairs ligated into the vector using standard method."
BASE COUNT          158 a 94 c 88 g 134 t 22 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 36; DB 17; Length 496;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 36
DB 18 CGGATCCACTACTAAGCGCGGCTACTGCTGTGAAT 53
RESULT 12
LOCUS AA514190
DEFINITION HFE1EST-741 Human fetal liver (S.Xue) Homo sapiens cDNA, mRNA
ACCESSION AA514190
VERSION AA514190.1 GI:2253714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartida; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 503)
TITLE Human liver ESTs
JOURNAL Unpublished (1997)
COMMENT Contact: Shepu Xue

```

Dept of Cellular Biology  
 Institute of Basic Medical Sciences/Beijing Union Medical College &  
 Chinese Academy of Medical Sciences (P.M.C. & C.A.M.S.)  
 5 Dong Dan San Tiao, Beijing, 100005 P.R. China  
 Tel: 8601-65296459  
 Fax: 8610-65240529  
 Email: xuosp@cdm.imicams.ac.cn  
 Seq primer: M13 Forward and Reverse Primers

Location/Qualifiers

## FEATURES

source

1. .503  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Human fetal liver (S.Xue)"  
 /tissue\_type="liver"  
 /dev\_stage="fetal"  
 /lab\_host="E.coli DH5a"  
 /note="Vector: pBluescript SK"  
 128 a 124 c 118 g 133 t

## BASE COUNT

ORIGIN

Query Match 1.2% Score 36; DB 9; Length 503;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 36

|||||

DB 24 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 59

## RESULT 14

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

Genome Structure and Immune Functions

Centre d'Immunologie INSERM-CNRS

Case 906, 13288 MAPSEILLE Cedex 9, FRANCE

Tel: 330491269496

Fax: 330491269430

Email: jordan@ciml.univ-mrs.fr

This sequence was determined at Genome Express, Grenoble, France

Seq primer: T7

Location/Qualifiers

1. .555

/organism="Mus musculus"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="MTA.C02.079"

/clone\_lib="MTA adult mouse thymus library"

/lab\_host="MC1061 p1"

/note="Vector: pCDNA3; Site 1: NotI; Site 2: EcoRI; The

cDNA library was constructed from poly(A)<sup>+</sup> RNA of an adult

mouse thymus by oligo dT primed reverse transcription.

cDNA was selected on gel for size above 800 nucleotides

after second strand synthesis then directionally cloned

into the pCDNA3 vector (NotI on polyA side, EcoRI on the

5'side)."

149 a 112 c 103 g 189 t 2 others

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 0;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 36

|||||

DB 15 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 50

## RESULT 14

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Wirtz MK

Glaucoma Genetics Lab

Oregon Health Sciences University

3375 S.W. Terwilliger Blvd., Portland, OR 97201-4147, USA

Tel: 503-494-4598

Fax: 503-494-6875

Email: wirtzmk@ohsu.edu

Seq primer: T7 Reverse

High quality sequence step: 350.

Location/Qualifiers

1. .612

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="104138"

/clone\_lib="Human Trabecular Meshwork cDNA library"

/tissue\_type="eye"

/cell\_type="trabecular meshwork"

/dev\_stage="2 week to 2 year old infants"

/lab\_host="Tc1010"

/note="Vector: pCDNA3; Site 1: EcoRI; Site 2: EcoRI; Human

cDNA library made from mRNA isolated from trabecular

meshwork cells established from eyes of 6 individuals,

ages 2 weeks to 2 years. Cells were harvested at passages

3 through 6. Invitrogen made a unidirectional cDNA library

from the mRNA from the frozen cells using a pCDNA3 vector

and Tc1010 host cells."

158 a 127 c 145 g 142 t 7 others

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 0;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 36

|||||

DB 9 CGGATCCACTAGTAAAGCGGCGGAGTGTGGTGGCAAT 44

## RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Wirtz MK

Glaucoma Genetics Lab

Oregon Health Sciences University

3375 S.W. Terwilliger Blvd., Portland, OR 97201-4147, USA

Tel: 503-494-4598

Fax: 503-494-6875

Email: wirtzmk@ohsu.edu

Seq primer: T7 Reverse

High quality sequence step: 350.

Location/Qualifiers

1. .612

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="104138"

/clone\_lib="Human Trabecular Meshwork cDNA library"

/tissue\_type="eye"

/cell\_type="trabecular meshwork"

/dev\_stage="2 week to 2 year old infants"

/lab\_host="Tc1010"

/note="Vector: pCDNA3; Site 1: EcoRI; Site 2: EcoRI; Human

cDNA library made from mRNA isolated from trabecular

meshwork cells established from eyes of 6 individuals,

ages 2 weeks to 2 years. Cells were harvested at passages

3 through 6. Invitrogen made a unidirectional cDNA library

from the mRNA from the frozen cells using a pCDNA3 vector

and Tc1010 host cells."

158 a 127 c 145 g 142 t 7 others

BASE COUNT

ORIGIN



SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 633)  
 AUTHORS Wirtz M K., Samples J R., Xu H., Severson T. and Acott T S.  
 TITLE Expression Profile and Genome Location of cDNA Clones from an  
 Infant Human Trabecular Meshwork Library  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Wirtz MK  
 Glaucoma Genetics Lab  
 Oregon Health Sciences University  
 3375 S.W. Terwilliger Blvd., Portland, OR 97201-4197, USA  
 Tel: 503-494-4698  
 Fax: 503-494-6875  
 Email: wirtzm@ohsu.edu  
 Seq primer: T7 Reverse.

FEATURES  
 source  
 1..633  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Human Trabecular Meshwork cDNA library"  
 /tissue\_type="eye"  
 /cell\_type="trabecular meshwork"  
 /dev\_stage="2 week to 2 year old infants"  
 /lab\_host="TOP10F"  
 /note="vector: pcDNA3; Site\_1: EcoRI; Site\_2: EcoRI; Human  
 cDNA library made from mRNA isolated from trabecular  
 meshwork cells established from eyes from 6 individuals,  
 ages 2 weeks to 2 years. Cells were harvested at passages  
 3 through 6. Invitrogen made a unidirectional cDNA library  
 from the mRNA from the frozen cells using a pcDNA3 vector  
 and TOP10F host cells."  
 BASE COUNT 161 a 125 c 165 g 179 t 3 others  
 ORIGIN

Query Match 1.2%; Score 36; DB 14; Length 633;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CGGATCCACTAGTAACGCCGCTAGTGTGGTGAAT 36  
 ||||||||||||||||||||||||||||||||||||  
 Db 39 CGGATCCACTAGTAACGCCGCTAGTGTGGTGAAT 74

Search completed: July 16, 2003, 05:27:05  
 Job time : 4170 secs









```

; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/434,797
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/09/482,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10/46
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,484
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 89
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(575)
; OTHER INFORMATION: n = A,T,C or G
US-09-933-797-89

Query Match      1.2%   Score 36;   DB 9;   Length 575;
Best local Similarity 100.0%;   Pred. No. 1.7e 07;
Matches 46;   Conservative 0;   Mismatches 0;   Indels 0;

QY 1 CGGATCCACTAGTAAAGGGGGGGGCGATGCTGGTGAAT 36
      |||||||
DB 40 CGGATCCACTAGTAAAGGGGGGGGCGATGCTGGTGAAT 65

RESULT 5
US-09-443-218-1
; Sequence 1; Application US/09443218
; Publication No. US20030094610A1
; GENERAL INFORMATION:
; APPLICANT: Lowenthal, John W
; APPLICANT: Johnson, Michael A
; APPLICANT: O'Neil, Terri E.
; TITLE OF INVENTION: No US20030094610A1-i Uses of Avian Interferon Gamma
; FILE REFERENCE: 48-95C
; CURRENT APPLICATION NUMBER: US/09/443,218
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 08/755,381
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT AU95/00114
; PRIOR FILING DATE: 1996-03-05
; PRIOR APPLICATION NUMBER: AU PN1542/95
; PRIOR FILING DATE: 1995-03-06
; PRIOR APPLICATION NUMBER: US 09/272,042
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Gallus sp. (chicken)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14)...(625)
; NAME/KEY: mat_prime
; LOCATION: (191)...(625)
; NAME/KEY: 3'UTR
; LOCATION: (626)...(1079)
; NAME/KEY: 5'UTR
; LOCATION: (1)...(133)
US-09-443-218-1

Query Match      1.1%   Score 35;   DB 9;   Length 1079;

```

```

Best Local Similarity 100.0%;   Pred. No. 5.9e-07;
Matches 35;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY 2 GGATCCACTAGTAAAGGGGGGGGCGATGCTGGTGAAT 36
      |||||||
DB 1 GGATCCACTAGTAAAGGGGGGGGCGATGCTGGTGAAT 35

RESULT 6
US-09-933-797-112
; Sequence 112; Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/09/442,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(331)
; OTHER INFORMATION: n = A,T,C or G
US-09-933-797-112

Query Match      1.0%   Score 32;   DB 9;   Length 331;
Best local Similarity 100.0%;   Pred. No. 1.8e-05;
Matches 32;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY 1 CGGATCCACTAGTAAAGGGGGGGGCGATGCTGGT 32
      |||||||
DB 50 CGGATCCACTAGTAAAGGGGGGGGCGATGCTGGT 81

RESULT 7
US-09-796-b94-6573/C
; Sequence 6573; Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04

```

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-796-692-6573

SEQ ID NO 6573  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-10-040-862-6573

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-796-692-6573

SEQ ID NO 6573  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-10-040-862-6573

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-796-692-6573

SEQ ID NO 6573  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-10-040-862-6573

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-10-040-862-6573

SEQ ID NO 6573  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-10-040-862-6573

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-834-975-712

SEQ ID NO 712  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-09-834-975-712

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-834-975-712

SEQ ID NO 712  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-09-834-975-712

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-834-975-712

SEQ ID NO 712  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-09-834-975-712

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-834-975-712

SEQ ID NO 712  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-09-834-975-712

Query Match 0.9%, Score 29, DB 9, Length 101,  
Best Local Similarity 100.0%, Pred No. 0.00053;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-09-834-975-712

SEQ ID NO 712  
LENGTH: 101  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (101)  
OTHER INFORMATION: u=A,T,C or G

US-09-834-975-712







100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000







CC trafficking of leucocytes between tissues and lymphatic vessels or to  
 CC affect migration in and out of the thymus.

XX Sequence 219 BP; 74 A; 54 C; 52 G; 39 T; 0 other;

Query Match 1.2%; Score 36; DB 19; Length 219;  
 Best Local Similarity 100.0%; Pred No 0.00013;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 36  
 DB 35 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 70

## RESULT 3

AA25511  
 ID AAX25511 standard; cDNA; 1288 BP.

XX AC AAX25511;

XX 02-AUG-1999 (first entry)

DE Human herpes simplex virus receptor B5T74 cDNA.

XX HSV receptor; B5T74; infection; diagnosis; therapy; vaccine;  
 KW antiviral; assay; ds.

XX Human herpes simplex virus.

XX Key Location/Qualifiers  
 FH 80..1204  
 FT /\*tag= a

XX W09920761 A2.

XX 29-APR-1999

XX 22-OCT-1998; 98W0-US2342.

XX 22 OCT 1997; 9705 095553;

XX (UNM) UNIV MICHIGAN.

XX Fuller AO, Li Q, McLaren NC, Perez A, Subramanian G;

XX WPI; 1999-402740/25.

XX P-PSDB; AAY05796.

XX Human herpes simplex virus receptor B5 and HVEM compositions

XX Claim 2; Page 67-68; 89pp; English.

XX This is the nucleotide sequence of B5T74 clone cDNA that codes for  
 CC novel human herpes simplex virus (HSV) receptor B5T74 or B5 (see  
 CC AAY05796), which appears to be a type II membrane-spanning, cell  
 CC surface protein. Transfection of eukaryotic cells (refractory to  
 CC infection with HSV) with a vector containing a nucleic acid  
 CC encoding the HSV receptor renders the host cells permissive to HSV  
 CC entry and replication. In the present invention, the combination  
 CC of a novel porcine cell model system which is refractory to HCV  
 CC entry, along with specific HSV B5 and/or HVEM (see AAY05797)  
 CC receptor proteins enables the development of assays for screening  
 CC antiviral compounds and therapeutics. The assays are useful for  
 CC detecting the ability of agents to inhibit HSV entry or spread  
 CC and provide for facile high-throughput screening of compounds  
 CC suspected to be able to inhibit such entry, e.g. compound  
 CC libraries, peptide libraries etc., to identify potential drug  
 CC candidates. The invention also provides a vaccine comprising  
 CC the HSV receptor, an immunogenic polypeptide or fragments of the  
 CC polypeptide.

XX Sequence 1288 BP; 409 A; 237 C; 285 G; 356 T; 1 other;

Query Match 1.2%; Score 36; DB 20; Length 1288;  
 Best Local Similarity 100.0%; Pred No 0.00026;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 36  
 DB 18 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 53

## RESULT 4

AAV10366

ID AAV10366 standard; cDNA; 1982 BP.

XX AC AAV10366;

XX 01-JUL-1998 (first entry)

XX Wheat low affinity cation transporter LCT1 gene.

XX Wheat; low affinity cation transporter; LCT1; heavy metal; alkali;  
 KW modulation; transgenic plant; contamination; soil; ss.

XX Triticum sp.

XX Key Location/Qualifiers  
 FH 137..1474  
 FT /\*tag= a  
 FT /product= "LCT1"

XX W09804700-A1.

XX 05-FEB-1998.

XX 28-JUL-1997; 97W0-US13250.

XX 29-JUL-1996; 96US 9022722.

XX (RESC) UNIV CALIFORNIA.

XX Antosiewicz DM, Clemens S, Schachtman DP, Schroeder JJ;

XX WPI; 1998-130689/12.

XX P-PSDB; AAW40204.

XX Plant low affinity cation transporter - used for modulating heavy on  
 PT alkali metal uptake

XX Claim 1; Page 24-25; 31pp; English.

XX The present sequence encodes wheat low affinity cation transporter  
 CC (LCT1). The present invention also describes a transgenic plant,  
 CC preferably a member of the genus Brassica, comprising an expression  
 CC cassette containing a plant promoter operably linked to a heterologous  
 CC LCT1 polynucleotide. The LCT1 polynucleotide can be used in an  
 CC expression cassette for modulating heavy metal or alkali metal uptake  
 CC in a plant. The transgenic plants are useful for removing heavy metals  
 CC from contaminated soils.

XX Sequence 1982 BP; 354 A; 611 C; 569 G; 448 T; 0 other;

Query Match 1.2%; Score 36; DB 19; Length 1982;  
 Best Local Similarity 100.0%; Pred No 0.00026;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 36  
 DB 18 CGGATGCTAGTAAAGGCGGCGGAGTGTGGTGGAAAT 53

## RESULT 5

AAT38068

ID AAT38068 standard; cDNA; 1079 BP.

XX









CC The invention relates to human polypeptides (AA172941-AA193841) and  
 CC the encoded proteins (AA00010-AA01910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 426 BP; 172 A; 60 C; 113 G; 65 T; 16 other;

Query Match 1.0%; Score 30; DB 22; Length 426;  
 Best Local Similarity 100.0%; Pred. No. 0.063;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3082 CTCATATAAATAAAAAAAAAAAAAAAAAAAAA 3111  
 Db 132 CTCATATAAATAAAAAAAAAAAAAAAAAAAAA 161

# RESULT 12

AAV69433  
 ID AAV69433 standard; cDNA; 1889 BP.

XX AAV69433;

XX 15-MAR-1999 (first entry)

XX Human ed4-6 cDNA #2.

XX Endothelial differentiation gene receptor homologues, human, HEDG;  
 KW EDG-6; detection; diagnosis; drug screening; treatment; inflammation;  
 KW disease; viral; bacterial; fungal infection; allergic response; injury;  
 KW hereditary disease; lymphoma; carcinoma; lymphoid; neuronal; cascade;  
 KW lymphocyte trafficking; leukocyte trafficking; signal transduction;  
 KW rheumatoid synovium; autoimmune disorder; ss.

XX Homo sapiens.

XX W09854062-A1.

XX 26 Nov 1998.

XX 22-MAY-1998; 98WO-CA00487.

XX 22-MAY-1997; 97US-0861747.

XX (ALLX) ALLELIX BIOPHARMACEUTICALS INC.

XX Muir DG, Vyas TB;

XX WPI; 1994-070147/06

XX New isolated human endothelial differentiation gene-6 receptor  
 PT homologue - used to develop products for treating v.g. infections.  
 PT allergic responses, trauma, hereditary diseases, lymphoma, carcinoma  
 PT or autoimmune disorders

XX Disclosure; Fig 2; 59pp; English.

XX This sequence encodes a novel human endothelial differentiation gene  
 CC (EDG)-6 receptor homologue designated HEDG. The encoded protein can be  
 CC used for detection, diagnosis and drug screening. The protein can also  
 CC be used to treat inflammation or diseases including viral, bacterial, or  
 CC fungal infections, allergic responses, mechanical injury associated with  
 CC trauma, hereditary diseases, lymphoma or carcinoma, or other conditions  
 CC which activate the genes of lymphoid or neuronal tissues. It can also be

CC used for treating problems involving excessive lymphocyte and leukocyte  
 CC trafficking. Inhibitors of HEDG are useful for controlling signal  
 CC transduction and signalling cascades in cells of the rheumatoid synovium.  
 CC The protein can also be used for treating autoimmune disorders such as  
 CC myasthenia gravis.

XX Sequence 1889 BP; 300 A; 575 C; 537 G; 416 T; 1 other;

Query Match 1.0%; Score 30; DB 20; Length 1889;  
 Best Local Similarity 100.0%; Pred. No. 0.062;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCGCATAGTAAAGGAGGAGCAGTGTGG 40  
 Db 50 CGGATCGCATAGTAAAGGAGGAGCAGTGTGG 79

# RESULT 13

AAA40303/C

ID AAA40303 standard; DNA; 48 BP.

XX AAA40308;

XX 11-SEP-2000 (first entry)

XX M. tuberculosis Cu, Zn Superoxide dismutase, SOD, gene fragment.

XX Cu, Zn superoxide dismutase; antibody; tuberculosis; enzyme;

XX superoxide radical; macrophage; sodC; bacterium detection; ds.

XX Mycobacterium tuberculosis.

XX W0200029017-A1.

XX 25-MAY-2000.

XX GY-N-V-1549, GYW-MS2472.

XX 13-NOV-1998; 98US-0108309.

XX (YUNG) YUNG SHIN PHARM INC CO LTD.

XX Lee FS, Wu CH;

XX WPI; 2000-38/619/33.

XX P-PSDB; AAY96264.

XX Novel monoclonal antibodies targeted to Mycobacterium tuberculosis  
 PT superoxide dismutase, useful for detecting tuberculosis infection.

XX Example 1; Page 14; 28pp; English.

XX The present sequence is a gene fragment of Mycobacterium tuberculosis  
 CC Copper, Zinc Superoxide dismutase, Cu, ZnSOD, superoxide dismutase,  
 CC catalyses the conversion of superoxide radicals, which are mutagenic,  
 CC into molecular oxygen and hydrogen peroxide. Cu, ZnSOD is thought to be  
 CC important for bacterial survival in macrophages. Macrophages engulf  
 CC bacteria, and set about destroying them by generating superoxide  
 CC radicals. The bacterium counteracts by releasing SOD, to neutralise the  
 CC superoxide radicals. The present sequence is added to the truncated  
 CC Cu, ZnSOD coding sequence, (AAA40303) to produce a full-length version.  
 CC The T of the stop codon of AAA40303 is mutated to an A, to produce a lys  
 CC codon. The present sequence was then added downstream of the lys codon.  
 CC Antibodies can be developed which bind specifically to the full length  
 CC protein. The antibodies are useful for detecting tuberculosis  
 CC infection in animals, and for detecting the presence of Mycobacterium  
 CC tuberculosis. The antibodies are also used to isolate the Mycobacterium  
 CC tuberculosis Cu, Zn SOD.

XX Sequence 48 BP; 10 A; 15 C; 13 G; 10 T; 0 other;

Query Match 0.9%; Score 29; DB 21; Length 48;  
 Best Local Similarity 100.0%; Pred No. 0.2;

Matches 29, Conservative 0, Mismatches 0, Indels 0, Gaps 0.

QY 1 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 29  
 DB 40 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 12

RESULT 14  
 ABA97964  
 ID ABA97964 standard; DNA: 49 BP.  
 AC ABA97964:  
 XX 25-APR-2002 (first entry)  
 XX LAB shuttle vector multiple cloning site sequence SEQ ID NO 7.  
 DE Lactic acid bacterium; shuttle vector; LAB: DNA vaccine;  
 XX multiple cloning site; ds.  
 KW Synthetic.  
 OS JP2001340090-A.  
 PN 11-DEC-2001.  
 PD 09 MAR-2001; 2081JP-0067675.  
 XX 26-MAY-2000; 2000TW-011025  
 XX (ANAR-) ANARATA BIOTECH CORP LTD.  
 PA WPI, 2002-135978/18.  
 DR A lactic acid bacterium shuttle vector useful as a DNA vaccine carrier  
 XX comprises a nucleic acid sequence encoding a protein related to  
 PT replication of the LAB plasmid and a non-antibiotic-resistant selective  
 PT marker -  
 XX Example 5, Fig 4, 24pp; Japanese.  
 PS The invention relates to a lactic acid bacterium (LAB) shuttle vector  
 XX comprising:  
 CC (a) LAB plasmid sequence containing the plus starting point of  
 CC replication and a nucleic acid sequence encoding a protein related to  
 CC replication of the LAB plasmid; and  
 CC (b) a non-antibiotic-resistant selective marker and its promoter  
 CC sequence. The lactic acid bacterium shuttle vector can be used as a DNA  
 CC vaccine carrier, in selecting transformed cells, drugs and foods. The  
 CC present sequence is that of a multiple cloning site of one of the LAB  
 CC shuttle vectors of the invention.  
 XX Sequence 49 BP; 10 A; 14 G; 11 T; 0 other;

Query Match 0.9%, Score 29; DB 24; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 29  
 DB 11 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 39

RESULT 15  
 AAQ89695/c  
 ID AAQ89695 standard; DNA: 94 BP.  
 XX AAQ89695;  
 AC AAQ89695;  
 XX 07-NOV-1995 (first entry)  
 DE Vector back-end oligonucleotide.  
 XX

KW Universal cloning vector for expression of heterologous genes or  
 XX back-end oligonucleotide; ss.  
 OS Synthetic.  
 PN WO9510620-A.  
 XX 20-APR-1995.  
 PD 14-OCT-1994; 94WO-US11719.  
 XX 15-OCT-1994; 94US-0136148.  
 XX (MLCW) MALLINKPODT VETERINARY INC.  
 XX McMullen JP, Synenki BM, Zook CA;  
 PI WPI; 1994-161805/21.  
 XX Universal cloning vector for expression of heterologous genes or  
 PT cDNA - allows high levels of expression without the need for  
 PT sequence modification  
 XX Claim 9; Fig.3; 40pp; English.  
 PS A universal vector comprises a P1 promoter, a ribosome binding site,  
 CC the first 30 bp of the delta-7 pig somatostatin gene, a 21 bp front-  
 CC end oligonucleotide (pref. the dimer comprising the complementary  
 CC strands given in AAQ89695-94; containing multiple start and restriction  
 CC enzyme sites, a back-end oligonucleotide (pref. the dimer  
 CC comprising the strands given in AAQ89695-96) containing at least 1  
 CC restriction enzyme site, a transcription terminator (pref. the dimer  
 CC comprising the strands given in AAQ89697-98) and a drug resistance  
 CC marker.  
 XX Sequence 94 BP; 20 A; 26 G; 20 C; 20 T; 0 other;  
 SV Query Match 0.9%, Score 29; DB 16; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 0.19;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 29  
 DB 82 CGGATCCACTAGTAAAGGAGGAGGAGTGTG 54

Search completed: July 16, 2003, 00:14:45  
 Job time : 647 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2003, 08:03:57, Search time 147 seconds  
(without alignments)  
6964.025 Million cell updates/sec

Title: US-10-053-192-2  
Perfect score: 3111  
Sequence: 1 cygatccactagtaacggcc ..aaataaaataaaataaaataaaataaa a1111

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMR seq.\*  
2: /cgn2\_6/ptodata/1/ina/5R.COMR seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMR seq.\*  
4: /cgn2\_6/ptodata/1/ina/6R.COMR seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTDUS.COMR seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DR	ID	Description
1	36	1.2	219	3	US-08-795-430-6	Sequence 5, Appli
2	36	1.2	219	4	US-08-510-133A-25	Sequence 25, App
3	36	1.2	219	4	US-08-585-895-25	Sequence 25, App
4	36	1.2	219	4	US-09-355-700-6	Sequence 6, Appli
5	36	1.2	219	4	US-08-601-132-25	Sequence 25, Appl
6	36	1.2	1982	2	US-08-900-148-1	Sequence 1, Appli
7	35	1.1	1079	3	US-08-765-381-1	Sequence 1, Appli
8	35	1.1	1302	1	US-08-525-697-1	Sequence 1, Appli
9	33	1.1	870	4	US-09-333-599-1	Sequence 1, Appli
10	30	1.0	1889	3	US-08-861-747-3	Sequence 3, Appli
11	29	0.9	115	2	US-08-454-557C-75	Sequence 75, Appl
12	29	0.9	115	2	US-08-340-426D-75	Sequence 75, Appl
13	29	0.9	115	2	US-08-450-673C-75	Sequence 75, Appl
14	29	0.9	115	5	PT-US95-1711A-75	Sequence 75, Appl
15	29	0.9	147	4	US-08-604-165-17	Sequence 17, Appl
16	29	0.9	147	4	US-08-734-054B-17	Sequence 17, Appl
17	29	0.9	151	4	US-09-359-361-7	Sequence 7, Appli
18	29	0.9	170	4	US-08-604-165-16	Sequence 16, Appl
19	29	0.9	170	4	US-08-734-054R-16	Sequence 16, Appl
20	29	0.9	371	4	US-09-404-879A-365	Sequence 365, App
21	29	0.9	392	4	US-09-404-879A-381	Sequence 381, App
22	29	0.9	444	4	US-09-404-879A-461	Sequence 461, App
23	29	0.9	396	4	US-09-404-879A-383	Sequence 383, App
24	29	0.9	404	1	US-08-594-031-154	Sequence 154, App
25	29	0.9	415	4	US-09-404-879A-365	Sequence 365, App
26	29	0.9	444	4	US-08-821-827C-15	Sequence 15, Appl
27	29	0.9	440	4	US-09-290-202R-15	Sequence 15, Appl

C 28	29	0.9	741	1	US-09-616-468A-2	Sequence 2, Appli
C 29	29	0.9	741	4	US-09-054-228-2	Sequence 2, Appli
C 30	29	0.9	741	4	US-08-818-655-2	Sequence 2, Appli
C 31	29	0.9	744	4	US-09-175-658B-24	Sequence 24, Appl
C 32	29	0.9	752	4	US-09-175-658B-23	Sequence 23, Appl
C 33	29	0.9	908	4	US-08-991-789A-46	Sequence 46, Appl
C 34	29	0.9	908	4	US-08-062-451-46	Sequence 46, Appl
C 35	29	0.9	908	4	US-09-594-451-46	Sequence 46, Appl
C 36	29	0.9	942	4	US-09-874-027-5	Sequence 5, Appli
C 37	29	0.9	1064	4	US-08-960-048-3	Sequence 3, Appli
C 38	29	0.9	1123	1	US-08-458-023B-3	Sequence 3, Appli
C 39	29	0.9	1186	4	US-09-374-888A-89	Sequence 89, Appl
C 40	29	0.9	1233	4	US-09-015-188-4	Sequence 4, Appli
C 41	29	0.9	1272	4	US-08-909-758-3	Sequence 3, Appli
C 42	29	0.9	1362	3	US-08-323-425-4	Sequence 4, Appli
C 43	29	0.9	1410	4	US-09-378-088A-81	Sequence 81, Appl
C 44	29	0.9	1590	4	US-08-908-758-4	Sequence 4, Appli
C 45	29	0.9	1654	4	US-09-216-909-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-795-430-6  
Sequence 5, Affiliation US/08795430  
Patent No. 6130071  
GENERAL INFORMATION:  
APPLICANT: Alitkov, Kari  
TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)  
TITLE OF INVENTOR: Protein and Gene, Mutants Thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08795430  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FI96/00427  
FILING DATE: 01-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/471,573  
FILING DATE: 28-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/401,122  
FILING DATE: 14-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/595,895  
FILING DATE: 12-JAN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/510,133  
FILING DATE: 01-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/340,011  
FILING DATE: 14-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
PETERFARLEY/PATENT NUMBER: 244,672,3691  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 219 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-795-440-6

Query Match 1.2% Score 36; DB 3; Length 219;  
 Best local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 CGGATTCCTACTAAGCGGGGCGGAGTGTCGTGCAAT 36  
 |||  
 DB 35 CGGATCCACTAGTAAAGCGGGGCGGAGTGTCGTGCAAT 70

## RESULT 2

US-08-510-133A-25  
 Sequence 25, Application US/08510133A

Patent No. 6221839

GENERAL INFORMATION:

APPLICANT: Alitalo, Kari

Joukov, Vladimir

TITLE OF INVENTION: Receptor Ligand

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/510.133A

FILING DATE: 01-Aug-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

REFERENCE/DOCKET NUMBER: 28113/32863

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 219 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-08-510-133A-25

Query Match 1.2% Score 36; DB 4; Length 219;  
 Best local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 CGGATTCCTACTAAGCGGGGCGGAGTGTCGTGCAAT 36  
 |||  
 DB 35 CGGATCCACTAGTAAAGCGGGGCGGAGTGTCGTGCAAT 70

## RESULT 3

US-08-585-895-25  
 Sequence 25, Application US/08585895  
 Patent No. 6245530  
 GENERAL INFORMATION:  
 APPLICANT: Alitalo, Kari  
 APPLICANT: Joukov, Vladimir  
 TITLE OF INVENTION: Receptor Ligand  
 NUMBER OF SEQUENCES: 35  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/585,895  
 FILING DATE:  
 CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.

REGISTRATION NUMBER: 38,153

REFERENCE/DOCKET NUMBER: 28113/33072

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 219 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-585-895-25

Query Match 1.2% Score 36; DB 4; Length 219;  
 Best local Similarity 100.0%; Pred. No. 2.8e-06;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 CGGATTCCTACTAAGCGGGGCGGAGTGTCGTGCAAT 36  
 |||  
 DB 35 CGGATCCACTAGTAAAGCGGGGCGGAGTGTCGTGCAAT 70

## RESULT 4

US-09-355-700-6

Sequence 6, Application US/09355700

Patent No. 6361946

GENERAL INFORMATION:

APPLICANT: Ludwig Institute for Cancer Research

Heinrich University Licensing

Alitalo, Kari (U.S. only)

Joukov, Vladimir (U.S. only)

TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)

Protein and Gene, Mutants Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 59

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/355,790  
FILING DATE: 05-NOV-1999  
CLASSIFICATION: Unkn-wd  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/795,430  
FILING DATE: 05-FEB-1997  
APPLICATION NUMBER: PCT/FI96/00427  
FILING DATE: 01-AUG-1996  
APPLICATION NUMBER: 08/671,573  
FILING DATE: 28-JUN-1996  
APPLICATION NUMBER: 08/601,132  
FILING DATE: 14-FEB-1996  
APPLICATION NUMBER: 08/585,895  
FILING DATE: 12-JAN-1996  
APPLICATION NUMBER: 08/510,133  
FILING DATE: 01-AUG-1995  
APPLICATION NUMBER: 08/340,011  
FILING DATE: 14-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28967/34140  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-355-700-6

Query Match 1.2%, Score 36, DB 4, Length 219,  
Best Local Similarity 100.0%, Pred. No. 2.8e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 CGGATTCATAGTAAAGGCGGAGGCGGCGGAAAT 46  
|||||  
DL 35 CGGATTCATAGTAAAGGCGGAGGCGGCGGAAAT 70

RESULT 5  
US-08-601-132-25  
Sequence 25, Application US/08/01142  
Patent No. 6403088  
GENERAL INFORMATION:  
APPLICANT: Alitalo, Kari  
APPLICANT: Joukov, Vladimir  
TITLE OF INVENTION: Receptor ligand  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-TOS/MS-TOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/601,132  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:

NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28967/34140  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-601-132-25  
Query Match 1.2%, Score 36, DB 4, Length 219,  
Best Local Similarity 100.0%, Pred. No. 2.8e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGATTCATAGTAAAGGCGGAGGCGGCGGAAAT 36  
|||||  
DL 35 CGGATTCATAGTAAAGGCGGAGGCGGCGGAAAT 70  
RESULT 6  
US-08-900-148-1  
Sequence 1, Application US/08/06148  
Patent No. 5965792  
GENERAL INFORMATION:  
APPLICANT: Schroeder, Julian I.  
APPLICANT: Antosiewicz, Banuta M.  
APPLICANT: Schachman, Daniel P.  
APPLICANT: Clemens, Stephan  
TITLE OF INVENTION: Nucleic Acids Encoding Metal Uptake  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-TOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,148  
FILING DATE: 28-JUL-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 69/022,722  
FILING DATE: 29-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 024070-08/000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0300  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1982 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 137..1471

OTHER INFORMATION: /product- "low affinity cation  
transporter 1 (LCT1)"  
US-08 900-148-1

Query Match 1.2% Score 36; DB 2; Length 1982;  
Best Local Similarity 100.0%; Pred. No. 2.3e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36  
|||||  
DB 18 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 53

## RESULT 7

US-08-765-381-1  
Sequence 1, Application US/08765381  
Patent No. 6083724  
GENERAL INFORMATION:  
APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic  
TITLE OF INVENTION: sequences encoding same  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully Scott Murphy and Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City, New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765.381  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PNL542/95  
FILING DATE: 06-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU96/00114  
FILING DATE: 05-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESSER, LEOPOLD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-516-742-4343  
TELEFAX: 1-516-742-4366  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Chicken (Gallus sp.)  
CELL TYPE: T-cell  
CELL LINE: CC8.1h  
IMMEDIATE SOURCE:  
LIBRARY: CC8.1h  
CLONE: ChfEN-gamma  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 134..625  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 191..625  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..133  
FEATURE:  
NAME/KEY: 3'UTR

LOCATION: 626..1079  
US-08-765-381-1

Query Match 1.1% Score 35; DB 3; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 6.6e-06;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36  
|||||  
DB 1 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 35

## RESULT 8

US-08-525-697-1  
Sequence 1, Application US/08525697  
Patent No. 5795764  
GENERAL INFORMATION:  
APPLICANT: Christgau, Stephan  
APPLICANT: Andersen, Lene N  
APPLICANT: Kauppinen, Sakari  
APPLICANT: Heldt-Hansen, Hans P  
APPLICANT: Dalboeoe, Henrik  
TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5795764; No. 5795764disk of No. 5795764th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/525.697  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/WORK# NUMBER: 4004.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-8123  
TELEFAX: 212-878-9655  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0486/93  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..1192  
US-08-525-697-1

Query Match 1.1% Score 35; DB 1; Length 1402;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 36  
|||||  
DB 1 GGATCCACTAGTAAACGGCGGAGTGTGTGGGAAT 35

## RESULT 9

US-09-333-599-1



```
Sequence 1, Application US/09333599
Patent No. 6245898
GENERAL INFORMATION:
APPLICANT: Testa, Jacqueline E.
APPLICANT: Quigley, James P.
APPLICANT: Seandel, Marco
TITLE OF INVENTION: MONOCLONAL ANTIBODIES THAT RECOGNIZE ANTIGENS
FILE REFERENCE: SUNY
CURRENT APPLICATION NUMBER: US/09/333,599
CURRENT FILING DATE: 1999-06-15
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 870
TYPE: DNA
ORGANISM: Homo sapiens
US-09-333-599-1

Query Match
Best Local Similarity 1.1%, Score 33, DB 4, Length 870,
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATCCACTATTAAGCGGCGGAGTGTGTGTGAAT 36
DB 1 ATCCACTAGTAAAGCGGCGGAGTGTGTGTGAAT 33

RESULT 10
US-08-861-747-3
Sequence 3, Application US/08861747
Patent No. 6020158
GENERAL INFORMATION:
APPLICANT: MUNROE, Donald G.
APPLICANT: VIAS, Tejal B.
TITLE OF INVENTION: A HUMAN EDG-6 PEPTIDE H-M-L-1
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikaido, Marmelstein, Murray & Gram LLP
STREET: 655 15th St., NW, Suite 330 - G Street Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/861,747
FILING DATE: 22-MAY-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jahns, Kristina M.
REGISTRATION NUMBER: 41,092
REFERENCE/POCKET NUMBER: PR074-7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-861-747-3

Query Match
Best Local Similarity 1.0%, Score 30; DB 3; Length 1889;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGTGA 30
DB 50 CGGATCCACTAGTAAAGCGGCGGAGTGTGTG 79

RESULT 11
US-08-454-557C-75/C
Sequence 75, Application US/08454557C
Patent No. 5830670
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,557C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/POCKET NUMBER: 0609,3840003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2500
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-454-557C-75

Query Match
Best Local Similarity 0.9%, Score 29; DB 2; Length 115;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGAGTGTGTG 29
DB 62 CGGATCCACTAGTAAAGCGGCGGAGTGTGTG 34

RESULT 12
US-08-340-426D-75/c
Sequence 75, Application US/08340426D
Patent No. 5948634
GENERAL INFORMATION:
APPLICANT: de la Monte, Suzanne
APPLICANT: Wands, Jack R.
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
TITLE OF INVENTION: of Alzheimer's Disease
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,426D  
 FILING DATE: 14-NOV-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ludwig, Steven R.  
 REGISTRATION NUMBER: 36,203  
 REFERENCE/DOCKET NUMBER: 0609,384,0002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2500  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 75:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 115 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: both  
 US-08-450-426D-75

Query Match 0.99; Score 29; DB 2; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 0.004;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 29  
 DB 62 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 34

RESULT 13  
 US-08-450-673C-75/c  
 Sequence 75, Application US/08450673C  
 Patent No. 5948888  
 GENERAL INFORMATION:  
 APPLICANT: de la Monte, Suzanne  
 TITLE OF INVENTION: Neutral Thread Protein Gene Expression and Detection  
 TITLE OF INVENTION: of Alzheimer's Disease  
 NUMBER OF SEQUENCES: 121  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 STREET: 1100 New York Avenue, Suite 600  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,673C  
 FILING DATE: 30-MAY-1995  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ludwig, Steven R.  
 REGISTRATION NUMBER: 36,204  
 REFERENCE/DOCKET NUMBER: 0609,384,0004  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2500  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 75:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 115 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: both  
 US-08-450-673C-75

Query Match 0.99; Score 29; DB 2; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 29  
 DB 62 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 34

RESULT 14  
 PCT-US95-17111A-75/c  
 Sequence 75, Application PC/TUS9517111A  
 GENERAL INFORMATION:  
 APPLICANT: de la Monte, Suzanne  
 TITLE OF INVENTION: Neutral Thread Protein Gene Expression and  
 TITLE OF INVENTION: Detection of Alzheimer's Disease  
 NUMBER OF SEQUENCES: 121  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 STREET: 1100 New York Avenue, Suite 600  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/17111A  
 FILING DATE:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/340,426  
 FILING DATE: 14-NOV-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ludwig, Steven R.  
 REGISTRATION NUMBER: 36,203  
 REFERENCE/DOCKET NUMBER: 0609,384,0002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2500  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 75:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 115 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: both

PCT-US95-17111A-75  
 Query Match 0.99; Score 29; DB 5; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 0.004;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 29  
 DB 62 CGGATCCACTAGTAAAGCGGCGGACAGTGTG 34

RESULT 15  
 US-08-604-165-17  
 Sequence 17, Application US/08604165  
 Patent No. 6287557  
 GENERAL INFORMATION:  
 APPLICANT: Boursnell, Michael E.G.  
 APPLICANT: Indlis, Stephen C.  
 TITLE OF INVENTION: Viral Preparations, Vectors,  
 TITLE OF INVENTION: Immunogens, and Vaccines  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111-4187  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/604,165  
 FILING DATE: 21-FEB-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9503395.7  
 FILING DATE: 21-FEB-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GR 9515557 8  
 FILING DATE: 28-JUL-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9503322.0  
 FILING DATE: 16-FEB-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Linda R. Judge  
 REGISTRATION NUMBER: 42,702  
 REFERENCE/DOCKET NUMBER: A-63282/MHD/LRJ  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 871-1989  
 TELEFAX: (415) 949-8711  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 147 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-604-165-17

Query Match 0.9%; Score 29; DB 4; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGATCCACTAGTACGGCGCGGACGCTG 29  
 ||||||||||||||||||||||||||||  
 Db 46 CGGATCCACTAGTACGGCGCGGACGCTG 74

Search completed: July 16, 2003, 05:30:02  
 Job time : 148 secs















541 GTATACATGAGACTAATCTGATGCAAAAATATGATCCAGACACTCTGACACACTAGAT 600  
541 GCTACACATGAGACTAATCTGATGCAAAAATATGATCCAGACACTCTGACACACTAGAT 600  
601 AAGGTAGACTACAGCAAAATATGAGTGTAACTTGGCAACTTCTCCACACACTATGAC 660  
601 AAGGTAGACTACAGCAAAATATGAGTGTAACTTGGCAACTTCTCTCCACACACTATGAC 660  
661 AGTCTGGAATATCTGCAATGAGTGTGATCTTCAATCTTGTATTAAGGCAACAAATAT 720  
661 AGTCTGGAATATCTGCAATGAGTGTGATCTTCAATCTTGTATTAAGGCAACAAATAT 720  
721 GTCTCTTTAAGATCCCTCTCTGTACCAAAAAAAGAAAGAAATCTTTTAA 780  
721 GTCTCTTTAAGATCCCTCTCTGTACCAAAAAAAGAAAGAAATCTTTTAA 780  
781 CACCTGCAAGTATGATGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840  
781 CACCTGCAAGTATGATGCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840  
841 AGCTTTGCAATCAGCAAAATATATATGTTGATAGAGGCAATTTAACTGATAT 900  
841 AGCTTTGCAATCAGCAAAATATATATGTTGATAGAGGCAATTTAACTGATAT 900  
901 GTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
901 GTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
961 AAGGAGATAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
961 AAGGAGATAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
1021 AAGTTTACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
1021 AAGTTTACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
1081 CACGTTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
1081 CACGTTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
1141 AAAAAAGTGAAGCAAGCTTTGAATGATGATGATGATGATGATGATGATGATGATG 1200  
1141 AAAAAAGTGAAGCAAGCTTTGAATGATGATGATGATGATGATGATGATGATGATG 1200  
1201 GCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260  
1201 GCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260  
1261 CTTTCAACTTCCGCACTGCTGTAAAGAAAAAAGATGCGACATCTATTTGTCACCTGA 1320  
1261 CTTTCAACTTCCGCACTGCTGTAAAGAAAAAAGATGCGACATCTATTTGTCACCTGA 1320  
1321 ATATTATGTAAGGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
1321 ATATTATGTAAGGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
1381 AAGTATGCTATGTAAGGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
1381 AAGTATGCTATGTAAGGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
1441 AATGTCGAAGCAAGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
1441 AATGTCGAAGCAAGCAATAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
1501 AATTTTCTTCT 1560  
1501 AATTTTCTTCT 1560  
1561 GTCTCTCTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
1561 GTCTCTCTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
1621 GAATTTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680

1621 GAATTTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
1681 CCACGAAATGATTTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATG 1740  
1681 CCACGAAATGATTTGAGCAAAATATGATGATGATGATGATGATGATGATGATGATG 1740  
1741 GAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
1741 GAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
1801 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
1801 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
1861 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920  
1861 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920  
1921 AGAAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
1921 AGAAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
1981 GAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
1981 GAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
2041 TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
2041 TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
2101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
2101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
2161 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
2161 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
2221 ACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
2221 ACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280  
2281 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
2281 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
2341 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400  
2341 AAGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400  
2401 ATCTTATAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
2401 ATCTTATAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
2461 GTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520  
2461 GTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520  
2521 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
2521 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2580  
2581 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640  
2581 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640  
2641 AATGAGGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700  
2641 AATGAGGCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700  
2701 ATGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2760  
2701 ATGAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2760

D6	2760	ATGTGACACGACGCGTGTGGGTATCTCTGACCTTTCACAAACCGTGTCCATGTTTGTG	2760
QY	2761	CTACTGACTGTATTTTGTAGSAGAGAAACAAAATGAATACCTGTCTACTATATCGCC	2820
D6	2761	CTACTGACTGTATTTTGTAGSAGAGAAACAAAATGAATACCTGTACTATATCGCC	2820
QY	2821	CAATCAATTAAGACATGGGGCTTTGGTGCTTTCATGATTACATAAGATGTTTGAG	2880
D6	2821	CAATCAATTAAGACATGGGGCTTTGGTGCTTTCATGATTACATAAGATGTTTGAG	2880
QY	2881	CAGACGAGCAATAGACACACACCATCCACAGTCTCTGCTGTATGACACCCCT	2940
D6	2881	CAGAGAGAGAAATAGAACCCACACCATCCACAGTCTCTGCTGTATGACACCCCT	2940
QY	2941	TCTCTCTTTATGTTTGTCATGTATGSAAGAAATACATGCCATAATCTAATGTTTAAAAAGT	3000
D6	2941	TCTCTCTTTATGTTTGTCATGTATGSAAGAAATACATGCCATAATCTAATGTTTAAAAAGT	3000
QY	3001	CACGTGGGTGCAGATCTAGAGCTTAAGTAACGAGTCTGGGGTTTCAAAATGTTATATGTT	3060
D6	3001	CACGTGGGTGCAGATCTAGAGCTTAAGTAACGAGTCTGGGGTTTCAAAATGTTATATGTT	3060
QY	3061	CCATAAAATGCAAAATAAACACCTCCATATATAAAAAATAAAAAATAAAAA 3111	
D6	3061	CCATAAAATGCAAAATAAACACCTCCATATATAAAAAATAAAAAATAAAAA 3111	

RESULT 4			
AX400072			
LOCUS	AX400072	2286 bp	DNA
DEFINITION	Sequence 243 from Patent WO0218424.		
ACCESSION	AX400072		
VERSION	AX400072.1	GI:21336495	
			FAT 06 JUN 2002
			linear

ORGANISM	REFERENCE
Homo sapiens	Tang, Y. T., Asundi, V., Zhou, P., Xu, A. J., Pei, F., Zhang, T., Wang, J. R., Zhao, Q. A., Wang, D., Liu, C., Drmanac, P. T., and Weinman, T. Nucleic acids and polypeptides patent: WO 0218424-A 243 07-MAR-2002;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

FEATURES	
Source	Location/Qualifiers
	1..2286
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
CDS	72..1715

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CA033516.1"
/db_xref="GI:21336496"
/translation="MDILIRGNPKPQLEPEFVAKVYIKTPAMIGTILNENYIMHTVFSRYNIWPGTALLSHSTIPPGEVYYSKILRSCTNTNTEAPVIVSEGTWAPDCKNIRSKAFYSLSHTIPDFTDNCILIMKGEQFVATSFNVIIRINPTLETTERVYRKYVAENLASHPIYDEAGVNLNGSTVIEKRIKVIIEKIPATPECKKQKSGPMWTEVPCSLPSRLSPSYHSFGVTENVIPLRQPKLLIKMATVATIPMWSACIAFHEEKETVTHIDORTROPVOKFYETDAMVPHVNAVEDGCIIVDAVDNSLXOYLYEALNDEGENSRSLSPVTLRPAPVPHVNRNVEGNILIKVASTATLKEEDSDVDCQPEPLEYPAFGAKDEDDGVYSIAVSTDPQKPEPLILDAKSFTEIARASVDVDMMDHGLGFTIDMMDIKRKAASEBDRASDCHGAPLT"

```

Query Match	25.8%	Score 804	Pr. 6	Length 2286
Local Similarity	70.0%	Prd. No. 4	1-164	
Matches 1112	Conservative	0	Mismatches 470	Indels 6
			Gaps 2	
128	GAGCGATGAGACAAATTTAACTGGAACAAAGAAAGACATCCAGGCCCATATAAAGCTG	187		
67	GAATTAATGATATAATTTGGTGAATATGAAAAAAGATGAGAGAGATTTGAGAGAGCCCA	126		

[illegible]













AF271298 2137 bp mRNA linear ROD 27 FEB 2001  
 LOCUS AF271298  
 DEFINITION Mus musculus beta-carotene 15,15'-dioxygenase mRNA, complete cds.  
 ACCESSION AF271298  
 VERSION AF271298.1 GI:11420969  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Gradata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2137)  
 REFERENCE Redmond, T. M., Gentileman, S., Duncan, T., Yu, S., Wiggert, B., Gantt, E. and Cunningham, F. X. Jr. Identification, expression, and substrate specificity of a mammalian beta-carotene 15,15'-dioxygenase  
 J Biol Chem 276 (9), 6560-6565 (2001)  
 JOURNAL MEDLINE 2125869  
 PUBMED 11092891  
 REFERENCE 2 (bases 1 to 2137)  
 Redmond, T. M. and Gentileman, S.  
 TITLE Identical  
 AUTHORS Direct Submission  
 JOURNAL Submitted (22-MAY-2000) LNCM, NCI, NIH, Bldg 6, Rm 439, 6 Center Dr., MSC 2740, Bethesda, MD 20892-2740, USA  
 FEATURES  
 source  
 1. 2137  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 198..1848  
 /codon\_start=1  
 /product="beta-carotene 15,15'-dioxygenase"  
 /protein\_id="AF33982.1"  
 /db\_xref="GI:11320970"  
 /translation="MELIFGQKKKQLEPVQAKVTGSLPAMLOCTLLRNGPQHTVGE  
 SKYHMFQGLALHSFSLPGEFVPSVYLSQVYIANIENKLVSEFPIMAVPLP  
 KNIFSKASYSHTIPDFTDNLINIMKGEDFYATETINIRKIDQLETEKVDY  
 PKYAVNLATSPHYDFAGNVLM;I;SVLEK;PIKRYVFIPLAIVLPSKKKSPVH  
 AEVCSISRLSPSYHSGVTEENVYFLEQPEKDIILKMATVAVRGVSNASCSE  
 DRECKTYIHIDQFRKRPVPTIKFTYIPVAVHNAVEKSTVPLQVLAHDSVLYK  
 FYLANLNKDEPKSKLVSPTLRRAVYLHYDKDAEVSNIYKVSSTTATLAKKDDH  
 VTCQPELVLEFETPRINATNGKPYRTFAEYVMSVPTKIKYHITIKSSIKWSE  
 ESCWPAELFVPTGAKDEDEGVIIISAVISTDPQTPETLILDKSETELARASVND  
 MHLDHGFLTPDAQMNNAVKLPALIEFVENSHPILPLAEELSHSEMTFAQHSNSEE  
 BASF COUNT 633 a 518 c 556 g 429 t 1 other s  
 ORIGIN  
 Query Match 24.98: Score 775.8; DB 10; Length 2137;  
 Best Local Similarity 68.88; P-val No 5.4e-158;  
 Matches 1097; Conservative 0; Mismatches 492; Indels 6; Gaps 2;

486 GACTACGAGACCTGCAAAAAACGCTTTCTAAATTTCTGTAATGTCACACG 545  
 Db  
 481 ATTTCGAGTGTACAGGAAACATGCTGAGAACATATATAAAGTGGATGATTAAT 540  
 Qy  
 546 ATCCCGSATTACAGAAATGCTGTATTAATATATATATATATATATATATAT 505  
 Db  
 541 GTTACGATGAGACTAACTTCTATGAAAAAATTAATGTAAGATCTGAGACATAG 609  
 Qy  
 606 GTACATGAGAGAGCAACATGATAGTAAATGACGCTTATATATATATATATAT 665  
 Db  
 601 AAGGTAGACTACAGCAAAATGATGACTTAACTGGCAATCTGCACCACTATAG 660  
 Qy  
 666 AAGGTGATATACCGCAAAATATGAGTAAATGAGTAAATGAGTAAATGAGTAA 725  
 Db  
 661 AGTCGTGAAATATATGCAATGAGGACTTCAATTTCTGTAATGAGAGCAATATA 720  
 Qy  
 726 GAGGTGAGAGGCTGTAATATAGGTAATGAGTAAATGAGTAAATGAGTAAAT 785  
 Db  
 721 GTTCCTTTAAAGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777  
 Qy  
 786 GTGATATTTAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 845  
 Db  
 778 AAACATGAG 847  
 Qy  
 846 AAGTAAAG 905  
 Db  
 838 CACAGCTTGGAATACAGCAAAATATATATGCTGCTGCTGCTGCTGCTGCTGCT 897  
 Qy  
 906 CAAATGAG 965  
 Db  
 898 ATGTAAG 957  
 Qy  
 966 ATCTCTAAATGAG 1025  
 Db  
 958 CATAAAG 1017  
 Qy  
 1026 GATAG 1085  
 Db  
 1018 AATGAG 1077  
 Qy  
 1086 AATGAG 1145  
 Db  
 1078 AATGAG 1137  
 Qy  
 1146 GGT 1205  
 Db  
 1138 TTAAG 1197  
 Qy  
 1206 GTGAG 1265  
 Db  
 1198 AATGAG 1257  
 Qy  
 1266 AATGAG 1325  
 Db  
 1258 AATGAG 1314  
 Qy  
 1326 AATGAG 1385  
 Db  
 1315 CTTGAG 1374  
 Qy  
 1386 AATGAG 1445  
 Db  
 1375 AATGAG 1434  
 Qy  
 1446 CCAATGAG 1505  
 Db  
 1435 AATGAG 1494  
 Qy  
 1506 AATGAG 1565  
 Db  
 1495 GATGAG 1554  
 Qy  
 1566 GATGAG 1625  
 Db

QY 1555 TGTATTGTGTCTGAGCAATTAAGACCCCTTCTACTATCAGTGGATGCTAAATCA 1614  
 DB 1636 GCGATGCTCTACGAGATCCCAAGAACCTGGCTTTTCTACTATCTGGATGCTAAAGT 1685  
 QY 1615 TTCAAGATTCGCTGAGCAATTAAGTAAATTAATTAATTAATTAATTAATTAATTA 1674  
 DB 1686 TTACGAACTGGCTGAGCTCTGTTGATGGGAAATAGCACTGGATCTTAAGGTCTTA 1745  
 QY 1675 TTATACCAAGATGATTTGGGCTGACAGCA 1709  
 DB 1746 TTATCCGATGAGACTGGATGAGTGAACCA 1780  
 RESULT 10  
 AB062912 2207 bp mRNA linear ROD 12 JUN 2001  
 LOCUS Rattus norvegicus mRNA for beta-carotene 15,15'-dioxygenase,  
 DEFINITION complete cds.  
 AB062912  
 VERSION AB062912 1 GI:14349293  
 KEYWORDS  
 SOURCE Rattus norvegicus intestine cDNA to mRNA.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (sites)  
 AUTHORS Takitani,K., Ban,R. and Tamai,H.  
 TITLE Regulation of beta-carotene 15,15'-dioxygenase in oxidative stress  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2207)  
 AUTHORS Takitani,K. and Tamai,H.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JUN-2001) Kimioka Takitani, Osaka Medical College,  
 Department of Pediatrics; 2-7 Daigakumachi, Takatsuki, Osaka  
 569-8686, Japan (E-mail:ped016@po.h.osaka-med.ac.jp,  
 Tel:81-726-811221(ex:2476), Fax:811-726-84-6554)  
 FEATURES  
 source  
 1..2207  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /tissue\_type="intestine"  
 387..2087  
 CDS  
 /codon\_start=1  
 /product="beta-carotene 15,15'-dioxygenase"  
 /protein\_id="BAB0807.1"  
 /db\_xref="GI:14349294"  
 /translation="MELIRGRNKKKQLEPRATVTSIPAMLOGTLLRNGPMHTVG  
 SKYHMFEDGIALHSFISIRGEVFSYKLGSTNYANIEANPIVSEGTAAVDPDC  
 KNIKSAFSLSHITPDFTDNCILINIKGGEYATTEINIRKIDPOTLLEKVDY  
 KRYAVNLATSHHYDAGVNLNGTSTADKGTIKVMPKIPATAGSKRKNPKIKH  
 SEVCSIPSRSLSPSYSHFVSEGVENVVLEDPKLDILKMATAMRKNVSAVSCYT  
 CKEDKTYIHIIDOKTRKPVPTKEVTPMVFHVNVAYEEDGLVDVIAVFNLSYCF  
 FYLANLKDEEKSRLSVPTLRFAPVPLHVDKDAEVSNLKVSSTATALKEDDH  
 VYQPEVLYEGLELPRIYNAHNSKPYPIYEAFAVQMSVPTKILKYDYLTAKSLKWE  
 ESCMPAPLFPVTPGAKDEDDGVLSALISTDPKIPRLIILIDASFTLAAASVDL  
 MHLDELFTPDGKNNAVKOTPAKTQEDENSDHPITGLTAAGLSHENDPTAGHGSKL  
 "

BASE COUNT 598 a 566 c 550 g 493 t  
 ORIGIN  
 Query Match 24.8% Score 770.4; DB 10; Length 2207;  
 Post local Similarity 68.5% Pred. No. 8.1e-157;  
 Matches 1094; Conservative 0; Mismatches 495; Indels 6; Gaps 2;

QY 240 GCGAGGATGAGACCAAAATAGGAGACCAATTAATTAATTAATTAATTAATTAATTAATTA 299  
 DB 444 GCGAGGATGAGACCAAAATAGGAGACCAATTAATTAATTAATTAATTAATTAATTAATTA 533  
 QY 400 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 459  
 DB 534 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 613  
 QY 360 TGCACATATACAGTGAATATAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 419  
 DB 614 GAGACATATACAGTGAATATAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 673  
 QY 420 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 479  
 DB 674 GCGCTACCGACAGCCCTGACCAAAAATCTTTTCGAAAGCTTTCTCTACTGCTGCTGACAG 733  
 QY 480 CATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539  
 DB 734 CATCCCGACATTCAGAGCAACAGCTGCGATTAACATGATGATGATGATGATGATGATGATG 793  
 QY 540 TGGTACAGAGTGAAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599  
 DB 794 TGGTACAGAGTGAAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 853  
 QY 600 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659  
 DB 854 GAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 913  
 QY 950 CAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719  
 DB 914 GAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 973  
 QY 720 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 779  
 DB 974 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1033  
 QY 777 TAAACACCTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836  
 DB 1034 GAAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1093  
 QY 837 GCAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896  
 DB 1094 GCAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153  
 QY 957 TCATAGAGAGATTAAGAGTGGTTTCACTTGTATGATGATGATGATGATGATGATGATGAT 1016  
 DB 1214 GTCGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1273  
 QY 1017 GCAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1076  
 DB 1274 AAGCAAAATTCACAGACAGAGTCAATGATGATGATGATGATGATGATGATGATGATGAT 1333  
 QY 1077 TGGGACAGTGTGTTTGAATAGTGGTCAACAGAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1136  
 DB 1334 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1393  
 QY 1137 GTTAAAAAATGAGCAAAATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1196  
 DB 1394 GCTGGGCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1453  
 QY 1197 GAAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1256  
 DB 1454 GAAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1513  
 QY 1257 GAAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1314  
 DB 1514 GAAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1573





[illegible][illegible]

D6		1166	GAAATGCAATGGCTTTGAGAACCGAGGTCTTCTTATAAATTCATTGGTGCTGCACAAATAT	1285
OY		1114	AATAGCGTTAGCATATGTTTACTTAATAAAAAACI-----GNCMAAGACTTTGAA	1166
D6		1226	CACAAACAACTTGAACAATTTCATCTTAATATATTTAGSAAAGCTTGGGGAAGGAGCTTAT	1285
OY		1165	GTCACACACAGAGCTTAACCCACATCCCAACCTGCAAGCAGCTTCTGTGAGCTCGAGAT	1226
D6		1286	CAGCTGCATTAATTCACAGACCACAAATCTTGGCTGGAGAGTTGTTTGGCTTAAATGTC	1347
OY		1225	GACACAGATGCAAAAATAGTTTTTAATTTACACAACTTC--AACCTCCGACTGCT	1286
D6		1346	AGTTTGAATGCCCCGAGAGGAGCACACCTGAGCTGATCTTCTATCTATCTTATGAGCTGT	1490
OY		1282	GTAAGAAGAAAAATGAGCAGCATCTATGTCACACTGCAATATTATATCT-----	1347
D6		1406	GTCACAAATAGGCTTATGTAACCACTCTGAGCTCTATATGAATAATCTACATAGGAGAGCTCA	1490
OY		1330	-----GAAGCATAAGAACTGCTGCTGTACACATATACATACAAAGGTAAGAAATAT	1490
D6		1466	GAAAAATAAAGAGGCTATCTGAAATTCCTATATATATATATATATATATATATATATATAT	1545
OY		1381	AATATATGCTATGCAACCAAAATGCCAGAGGACCTCTCTACAAAAATATGCAAACTG	1441
D6		1526	TATCATTTCTTTTATGCGCTGTGCTTTGGCATTTAGTGGGAGATTCTCGATCAAGGTT	1585
OY		1441	AATGTCCAAATAAAGTAAAT	1500
D6		1586	GATGTGCTGAATAACACACTGAAGGTTTGAGAGAAAGAATGGCTTTATCCGACAAACCT	1642
OY		1501	ATCTTTGTTCCCAATCCCTGGATATTAAGAAACATATTAAGCTGTGTTTGACCTGTGTT	1566
D6		1646	GTTTGTCTTCCACACATTAAGCAATATTAAGAAACATATTAAGCTGTGTTTGACCTGTGTT	1705
OY		1561	GTGCTGCTTAAGCTAAATATAAACAACTTCTATATATATATATATATATATATATATATAT	1620
D6		1706	ATCAGTCCCAACCAAGATATGAAGAATTTTCTGCTACTTTGGATGCCAAGAACTTGAA	1765
OY		1621	GAATGAGCGCGACCTACATTAATATGAATAATATATTAAGCTGTGTTTGACCTGTGTT	1686
D6		1766	GAGTGTGCTGAGCAATGAAGCTTAATCTGTGTGATGATGATATATATATATATATATATAT	1825
OY		1681	CC 1682	
D6		1826	CC 1827	
RESULT 14				
AF276432		2597 bp	mRNA linear	PRI 16- AUG-2001
Locus				
DEFINITION			Homo sapiens putative carotene dioxygenase mRNA, complete cds.	
ACCESSION			AF276432	
VERSION			AF276432.1	GI:14582264
KEYWORDS			.	
SOURCE			Homo sapiens.	
ORGANISM			Homo sapiens	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
FEATURES			1 (bases 1 to 2597)	
AUTHORS			Cunningham,F.X.	
TITLE			An ortholog of the human retinal pigment epithelium protein RPE65	
JOURNAL			Unpublished	
REFERENCE			2 (bases 1 to 2597)	
AUTHORS			Cunningham,F.X.	
TITLE			Direct Submission	
JOURNAL			Submitted (08-JUN-2003) Cell Biology and Molecular Genetics,	
			University of Maryland, Microbiology Building, Campus Drive,	
			College Park, MD 20742, USA	
			Location/Qualifiers	
			1..2597	
FEATURES			/organism="Homo sapiens"	
source			/db_xref="taxon:9606"	
			/chromosome="11"	

[illegible]

























REFERENCE	Mammalia: Eutheria; Primates: Catarrhini Hominoidea; Homo.
AUTHORS	I (bases 1 to 592)
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.R.
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery
MEDLINE	Genome Res. 6 (9), 791-806 (1996)
COMMENT	97044477 Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu Tissue Procurement: Dr. Gregg Hageman CDNA library preparation: Dr M. Bento Soares, University of Iowa CDNA library Arrayed by: Dr. M. Bento Soares, university of Iowa DNA Sequencing by: Dr. M. Bento Soares, university of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) Seq primer: MJ3 Reverse.  FEATURES SOURCE 1..592 Location/Qualifiers organism="Homo sapiens" db_xref="taxon:9606" clone="U1-E-EJ0-aln-n-09-0-U1" cdate="U1-E-EJ0" library_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Pooled and Macular, RPE and Choroid" dev_stage="fetal and adult" lab_host="DH10B (Life Technologies) (TI phage resistent)" note="Organ Eye: Wistar-Kyoto pr/TTK-Pao (Pharmacia) with a modified polylinker Site_1, EcoR I, Site_2, Not I, U1-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTTK-pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dl)II tail. This sequence tags for this library are: fetal eyes, AGAATGACAGCA ; lens, CGATTAGCCA; eye anterior segment, AATGGCCCATC ; optic nerve, CTATTAACTG; retina, GCGTC; Retina Pooled and Macular, GTGCC; RPE and Choroid, ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT	161 a     164 c     141 g     123 t     3 others
ORIGIN	
Query Match	8 4% Score 261 f; DB 14 Length 592;
Best Local Similarity	66.2%, Fred No. 2, C=3%
Matches 391:	Conservative 0; Mismatches 197; Indels 3; Gaps 1;
CY	938 GGSCCTTCCTGCCTTGCTTGTATAAGSGAGTAAGAAGCTGTTGACTTTAGACACA 997
Dd	1 GGSACTTCCTTCTGCTTTTCTCACAGGAGGAGAAATTAATATWATADTDAATAA 60
CY	938 ACACAAAAAACGACTATATATTAAGTTTATATGATATTTTGTATATATATATA 10572
Dd	61 GTAATGAGTAGCTTTTGTACASATTAAGTTTTACAAAAGGAGGAGGAGGAGTACC 120
CY	1058 TAAATGCTTACAAAGACATGGCCACGTTGTTTGTATATCGTTGCTTACAGACATA 11177
Dd	121 TCAGAGCTATCAAAGAGATGAGTATATGTTTGTTTATATATATATATATATAT 180
CY	1118 GCTTTAGCATATGTTTATTAAAAAAACAGGACAGCAAGCTTTTAAAGTGACACAGC 11777
Dd	181 GCCTTACAGAGTTCTTTACTTGGCCAACTGAAACGACAGAGACTTAAAGAGAACTGC 240
CY	1178 TTACTCTATCTCAATGCAAGGAGTTTGTGTCCTCTGTATATATATATATATATAT 1277







```
Db      643 AAAAATTTACAGTGGCTACAGCATTATTAGATGCT-----CCACTGAAACCA 693
QY      760 AAGAGAGAAATCTTGTTTTAAACACGCTGAGTAGATGCTGCAACCTCTGGCTCCG 819
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      694 GACAGAGAGATCTCAATAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 753
QY      820 CTCGACGACAGCTACTAGCAGAGCTTTGGAGATGACAGAGAGAGAGAGAGAGAGAG 879
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      754 TTTAAGGCAATTTATGTAACAGATTTGATCTGACCTGCAACCTATATGCTTTGAG 813
QY      880 GAGCCATTTAACTGGATATGCA 905
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      814 ACTCCAGTCAAAATTAACCTTTCAA 839
```

Search completed: July 15, 2003, 21:28:40  
Job time : 4173 secs





QY	51	CATTATGAAACGCGGGAGGATATTATATGTGTAACGTAGAGCTAACTTCATCGAAAAAT	572
DB	432	CGGTACACCCAGTAGGGAGGATTTATATAGGCTGGAGGGAACTTCATTACAAAGAT	491
QY	573	TGATTCAGCAACTCTGGAGACTCTGATGATAGGGTAGAGTACAGCAATATATAGCTAA	632
DB	492	TAACTCTGAGACCTCGGAGTAAATTAAGAGAGTTGATCTGTGAACGTAGCTGTCTGCA	551
QY	633	CTTGCAACTTCTACCTACACTATGACATGTGTGAAATATCTTAAATATGGTACTTC	692
DB	552	TGTAAGCTAACTGTACAGCTCAATTAAGAAATGATGGAGTCTTAAAGTCTTCCTGAC	711
QY	693	AATGTGATTAAGGAGCAACAAATATGTTCTCTTAAGAGCTTCCTTCCTGACGAGA	752
DB	612	CTTGGGAAAAATTTTTCATTGCTTCCTACATATATTTGAAAGATCTT-----GTACT	662
QY	753	AAAAAGAAAGCAAACTCTGTTTAAATACCTGGAAATGATATCTATATCTGTCTG	812
DB	663	CTAAAGAGAGTAAGAGATGATCAATTAAGCAAGCTCGAGCTGTCTGTGCAATTCCTGCA	722
QY	813	CTCTCTCTGTCGAACCTAACTTATCTACCTAGCTCTTGGGAAATGCAAAATATATCTG	872
DB	723	CGACTCTATTAAGCCATCTCTACCTGTCTATAGTTTGGTTTACCTCTCACTATATCTGT	782
QY	873	CATAGAGCACTATTAACTGGATATGTAAATTTGTAATCTGCTTAACTCTGCTAACT	929
DB	783	TGCTGTAAGCTCAAGTAAATTAATCTCTGTAAGCTCTCTCTGTCTGTCTGTCTG	842
QY	930	TCTGCACTGGGCT	989
DB	843	AGCTCACTATCATGATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	902
QY	990	AATCAAAATATGAAAAAGAGATGCAACCAAGTCTTAACTGATCTGTGTGTGT	104
DB	903	TGCAAAAAAAATAAAAAGTATCTGATTAATAGATATCTCTCTCTCTCTCTCTCT	962
QY	1050	TCACCCATTAATGCTTAGAAGAGATGGCCAGCTGTTTGTGATCTCTGCT---A	1104
DB	963	CTATCTATCAATACCTAGCAAGAAATATGATCTCTATTTGAGATCTGTCTGTGGA	1022
QY	1107	CATAGCAATATGCTGTACGATGCTTTTACTTAAAAAAGCTGCAAAAGCTTTGAT	1166
DB	1023	AGCATTTGCAATGCTGTCAATTTCTGATTTATTTAGCCATTTAGTCAAGACCTGGAA	108
QY	1167	GAAATCACTTACTCTCTATCTCAATCTGTAAGCTCTT-----GTCTGCTCTGCA	1222
DB	1083	CGTGAAGAAAAATGCGAGAAAGGCTGCGCAGCTGAGATAGAGATCTCTCTCT	1144
QY	1221	GTTGTCAGAGATGCGAAGTATGTTTAAATTTGATCAAACTGCT---AATCTGCGCAG	127
DB	1143	GAAATTCGAAAGCTGTGAATATGATTAATCTGATTAATCTGATATGCTGCAAGAGG	120
QY	1278	TGCTGTAAAGAAAAAGATGCGAGCATCTATTGTCAACTGAAATATTAT-----	1322
DB	1203	TCTAACTCTGCTATGAGAGAGCATCTGCTGGAAGCTGAGCTCTCTCTCTCTCT	1266
QY	1328	GTTGAGGATAGAACTGCTCTGCTGTAATCTATGATCAATAGGCAAAATATCAAGTA	1388
DB	1263	TGCTTAAGCTTTGACTTCTCTTAATTAATCTATGAGAGATATGCTGGGAGGCTTCA	1322
QY	1486	TGCTATGTCAGAGAGTCACTGAGAGGAGTCTGATCAAAAGATGCAAACTGATCT	1444
DB	1323	GTATGCTATGAGAGCTGCTGATCAATCTGCTGAGAGGAGGCTGTAAGCTGAG	1388
QY	1446	CTAAACAAAGCAAGTACTCTATGAGGAGGAGGATCTCTCTCTCTCTCTCTCTCT	1502
DB	1383	CAGAGCTAAAGCAAGCTGATATGCTGAGAGAGCTGATCTATCTGATCAAACTGATCT	1444
QY	1506	TCTTCTGCACTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1566
DB	1443	TCTTCTGCACTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1502

[illegible]



[illegible]

```

1 Patent No. 5579393
2 GENERAL INFORMATION:
3 APPLICANT: R. W. Class Clot Eriksson, UIF
4 TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which
5 TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding
6 Patent No. 5579393
7 TITLE OF INVENTION: Therefor, And Uses Thereof
8 NUMBER OF SEQUENCES: 5
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Felte & Lynch
11 STREET: 805 Third Avenue
12 CITY: New York City
13 STATE: New York
14 COUNTRY: USA
15 ZIP: 10022
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
18 OPERATING SYSTEM: PC-DOS
19 SOFTWARE: MultiPerfect
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/04/200,407
22 FILING DATE:
23 CLASSIFICATION: 435
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 07,444,549
26 FILING DATE: 15-MAY-1992
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Hanson, No. 557939man D.
29 REGISTRATION NUMBER: 40,946
30 REFERENCE/DOKKT NUMBER: int 280
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (212) 688-9200
33 TELEFAX: (212) 848-3884
34 INFORMATION FOR SEQ ID NO: 1:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 2629 bases
37 TYPE: nucleic acid
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: cDNA to mRNA
41 HYPOTHETICAL: no
42 ANTI-SENSE: no
43 US-OR-200-807-1
44
45 Query Match 7.9% Score 247 DB 1 Length 2629:
46 Best Local Similarity 51.0% Pred. No. 4.5e-56:
47 Matches 741, Conservative 0, Mismatches 715, Totals 45, Gaps 7:
48
49 1bu GAAGAAATCCGACAGTCCCAAAAACATGAGCGGGGAAGCTCAGTTGGATTGGTCGA 219
50 71 GAATAAATATCATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 130
51 220 AGGCAATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 279
52 131 GCAATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 190
53 280 TGGTTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 349
54 191 ATTTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 250
55 340 ACAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 399
56 251 CAACAAGAGGTATACAGACACATCAATCAATCAATCAATCAATCAATCAATCAAT 310
57 400 GTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 459
58 311 ATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 370
59 460 TTCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 519
60 371 TTCTTACTT-----CCGAGAGTAGAGTTACTGCACATGCCCTTGTAATATCTAC 424

```









US-09-124 698 28

## Query Match

1.3%, Score 41.8; DB 3; Length 2307;

Best Local Similarity 51.3%; Pred. No. 0.22;

Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

DB

1667 ATGGGATGTTTATACGAGATATTTGCGGCTGAGACGAGAAATAAAGCTATTCATC 1726

DB

2153 AAGGAGTGTTCACCAAAAGGCGGCTTGGGAGAGACACACACAAAGCTATTAATAA 2094

DB

1727 GCAGTACGACAAAGTACGACACTTTCTACGACATGAGTTAATATCCCTTTACGATTC 1786

DB

2093 CGATATCTTCTGCTGCAAAACCTATGAGAAATAAATTTAAAGATATAAAACCTG 2034

DB

1787 AAGAACACCATATATACGACACAAATGACTATGATATCTTTAATATATACATATAA 1846

DB

2033 ACCAAGACCATATATATACAAACCAATATATATTAATCTGCTTCTATGATATAT 1974

DB

1847 TCCCTTTAA 1855

DB

1974 GTCAAGCTTA 1965

## RESULT 11

US-09-127-480-28/c

Sequence 28, Application US/09127480

Patent No. 6194151

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESS: TESTA, HORWITZ &amp; THIBEAULT

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

FILING DATE: US/09/127,480

CLASSIFICATION:

PRIORITY APPLICATION DATA:

FILING DATE: US/08/592,541

ATTORNEY/AGENT INFORMATION:

NAME: Pilsch, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 2307 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-127-480-28

Query Match

1.3%, Score 41.8; DB 4; Length 2307;

Best Local Similarity 51.3%; Pred. No. 0.22;

Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

DB

1667 ATGGGATGTTTATACGAGATATTTGCGGCTGAGACGAGAAATAAAGCTATTCATC 1726

DB

2153 AAGGAGTGTTCACCAAAAGGCGGCTTGGGAGAGACACACACAAAGCTATTAATAA 2094

DB

1727 GCAGTACGACAAAGTACGACACTTTCTACGACATGAGTTAATATCCCTTTACGATTC 1786

DB

2093 CGATATCTTCTGCTGCAAAACCTATGAGAAATAAATTTAAAGATATAAAGCTG 2034

DB

1787 AAGAACACCATATATACGACACAAATGACTATGATATCTTTAATATATACATATAA 1846

DB

2033 ACCAAGACCATATATATACAAACCAATATATATTAATCTGCTTCTATGATATAT 1974

DB

1847 TCCCTTTAA 1855

DB

1974 GTCAAGCTTA 1965

RESULT 12

US-08-496-841C-28/c

Sequence 28, Application US/08496841C

Patent No. 6210919

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMMENS, JOHANNA M

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESS: Parby &amp; Parby, PC

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

FILING DATE: US/08/496,841C

CLASSIFICATION: &lt;unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Paul F. Feilner, Ph.D.

REGISTRATION NUMBER: 35,135

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 527-7700

TELEFAX: (212) 753-4237

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 2307 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 28:

US-08-496-841C-28

Query Match

1.3%, Score 41.8; DB 4; Length 2307;

Best Local Similarity 51.3%; Pred. No. 0.22;

Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

DB

1667 ATGGGATGTTTATACGAGATATTTGCGGCTGAGACGAGAAATAAAGCTATTCATC 1726

DB

2153 AAGGAGTGTTCACCAAAAGGCGGCTTGGGAGAGACACACACAAAGCTATTAATAA 2094

DB

1727 GCAGTACGACAAAGTACGACACTTTCTACGACATGAGTTAATATCCCTTTACGATTC 1786

DB

2093 CGATATCTTCTGCTGCAAAACCTATGAGAAATAAATTTAAAGATATAAAGCTG 2034

DB

1787 AAGAACACCATATATACGACACAAATGACTATGATATCTTTAATATATACATATAA 1846

DB

2033 ACCAAGACCATATATATACAAACCAATATATATTAATCTGCTTCTATGATATAT 1974

















```

; NAME/KEY: misc_feature
; LOCATION: (1)...(478)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-21852

```

Query Match	1.48;	Score 42.8;	DB 9;	Length 478;
Best Local Similarity	67.0%;	Prod No. 0.93;		
Matches	59;	Conservative	0;	Mismatches 29;
				Indels 0;
				Gaps 0;

[illegible]

```
QY      3084 CCATATAAAAAAAAAAAAAAAAAA 3111
          |  ||| ||||| ||||| |||||
Db      388 CNAAAAAAAAAAAAAAAAAA 415
```

RESULT 10  
US-10-239-676-16  
Sequence 16 - Amplification (US/10239676)

```

; GENERAL INFORMATION:
;
; APPLICANT: OLEK, Alexander
;
; APPLICANT: PIEPENBROCK, Christian

```

```

:
:
: TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
:
: FILE REFERENCE: 5013.1003
:
: CURRENT APPLICATION NUMBER: US/10/2339, 676
:

```

PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8

PRIOR FILING DATE: 2001-04-06  
2000-04-06

2000-09-01  
NUMBER OF SEQ ID NOS: 228

TYPE: DNA  
ORGANISM: Artificial Sequence

```

FEATURE:      UNKNOWN
NAME/KEY:     unsure
COMMENT:      none

```

Query Match	Score	DB	Length
1.48	42.2	9	6046

QY 2352 AATGCTTTATGTACATTCATTACATTACTCAGTGTCTATTCACATCCCTAATGTTAAGC 2411

2412 AATT...GATTAAATCAATTTTCCTTGAGAAATCTGACAAACAAATTTGTGTAATG 2465

2463 AGATGACATGTTGTTAAAGATGAACAGGAATATCTTAAIACATATGCTTATTTCT 2528

2529 GTTACTAATGATGATATGCAATGAGACCAATGTATTTCAGAGAACTCAGATATACA 2588

UY 2589 TTCACCAATTCTGTAGGTGMAAATGCATT 2619

Db 3053 TGGTAGTATTAGCTTATTAAATTATTTT 3083

RESULT 11  
US-10-239-676-139/c

```
; GENERAL INFORMATION:
; APPLICANT: OLFK, Alexander
; ADDITIONAL INFORMATION: 01-14-1988
```

1 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
2  
3 FILE REFERENCE: 5013-1003  
4  
5 CURRENT APPLICATION NUMBER: US 10/8720 676

PRIOR APPLICATION NUMBER: PEI/HOI/03968  
DE 10019058.8  
DE 10019173.8

```

; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06

```

NUMBER OF SEQ ID NOS: 228

```

; LENGTH: 8866
; TYPE: UNA
; ORGANISM: Artificial Sequence

```

```

1 OTHER INFORMATION: chromosomally located genomic DNA (Homo sapiens)
2 FEATURE:
3 NAME/KEY: unsure
4
```

```

LOCATION:
NAME/KEY: unsure
LOCATION: (5595, 5601)

```

Query Match	1.48;	Score 42.2;	DB 9;	Length 8866;
Best Local Similarity	45.78;	Pred. No. 9.6;		

QY 2179 ATTGATACCAATGCAATTTCTTGATGGTGGCAATTTACAGTAATACCAATTAATTAAT\* 22+4

QY 2239 ATTACGCTACCACTTTCTAATTACACCGAAACAAACAGAGAAACAGATTGCT 2249

2289 TTTTATATGCGCATCTGTGTAATTTTGTCTTTCATGCA/AAAAAAGATTA/ATATATCT 2408

[illegible]

QY 2419 ATT-----TAATCAGCTTTTCCTGAGCAATATCTGACATACAGTTTGTGTAATGAAAG 2418

2474 ACATATGTTCTCTAAAGATGAAACAGGAATGATCTTTAATAGTATTGTTAATGTGTAC 2514

Db 1962 TCACATCAACGACATATAAAATTTCTATCTATAAAATCTATTAAATTAAATTAAATTAACTTTCTC 190

```

RESULT 13
US-09-938-842A-3170/G
: Sequence 3170, Application US/09-938842A
: Patent No. US66201657B1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Karpis, Joel
: APPLICANT: Wang, Xun
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT300-3
: CURRENT APPLICATION NUMBER: US/09/938-842A
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16

```

[illegible]

Accession	Sequence	Position
D6	2522 CCGAGATCCCGCATCCCGCCACAAACATCATGGAGACAAAATTTTAAGCATTAACCAAAACCC	2463
QY	1787 AACAAACCGCATATACGACACAAAATGACTATGATATCTTTAAATAAATAGATATATA	1846
D6	2462 ACCAAATACCGCATATTACAAACCAATATGCTAACTGTGTCCCTTTCTATGATATATAT	2403
QY	1847 TCCCTTTAA	1855
D6	2402 GTATCTGTA	2394

```

1  RESULT:15
2  US-10-038-001-1
3  : Sequence 1, Application US/10038001
4  : Publication No. US20020187952A1
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: PALMER, Kenneth E.
9  : APPLICANT: POORE, Gregory P.
10 : APPLICANT: MCCORMICK, Allison
11 :
12 : TITLE OF INVENTION: ROLLING CIRCLE REPLICAT
13 :
14 : TITLE OF INVENTION: VECTORS
15 :
16 : FILE REFERENCE: 008010179C:PU501
17 : CURRENT APPLICATION NUMBER: US/10/038,001D
18 :
19 : PRIORITY FILING DATE: 2001-12-20
20 : PRIOR APPLICATION NUMBER: 09/505,477
21 :
22 : PRIOR FILING DATE: 2000-02-16
23 :
24 : NUMBER OF SEQ ID NOS: 9
25 :
26 : SOFTWARE: FastSeq for Windows Version 4.0
27 :
28 : SEQ ID NO: 1
29 :
30 : LENGTH: 5285
31 :
32 : TYPE: DNA
33 :
34 : ORGANISM: Porcine circovirus
35 :
36 : US-10-038-001-1

```

Query Match	1.38;	Score 41.8;	DB 9;	Length 5285;
Best local Similarity	75.48;	Pred. No. 8.6;		
Matches 52; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

[illegible]

Search completed: July 16, 2003, 00:03:50  
Job time : 490 secs





beta, beta-carotene 15,15'-dioxygenase protein, nucleic acids and antibodies, useful for production of vitamin A from carotene and gene therapy of ophthalmological disorders

Claim 1: Page 12-13; 37pp; English.

Beta, beta-carotene-15,15'-dioxygenase (bcoD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bcoD are useful as primers of probes for specific amplification and/or detection of the gene that encodes bcoD, for isolation of related sequences in other organisms, for determining bcoD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bcoD gene. The nucleotide encoding bcoD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bcoD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bcoD are used for detection/quantification of bcoD in immunoassays.

Sequence 526 AA;

Query Match 100.0%; Score 526; DR 21; Length 526;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

0Y 1 METFFNNKKEHPPIKAEGOLPTWLOGVLLRNGPCMH1IGDTKYNNHWPGLALHSF 60
Db 1 METFFNNKKEHPPIKAEGOLPTWLOGVLLRNGPCMH1IGDTKYNNHWPGLALHSF 60
0Y 61 TPKNFVYVYSKYI PSHYVNCNIEANPIVVSFEGTMAVDPCKNIFAKAFSYLSHTPEF 120
Db 61 TPKNFVYVYSKYI PSHYVNCNIEANPIVVSFEGTMAVDPCKNIFAKAFSYLSHTPEF 120
0Y 121 TDNGLINIMKTGDDYVATSETFNTRKIDPOTLETLEKVNYSKVAAVNA1AISHRHDSACN 180
Db 121 TDNGLINIMKTGDDYVATSETFNTRKIDPOTLETLEKVNYSKVAAVNA1AISHRHDSACN 180
0Y 181 I1NNGTSIVDKGRTKTVLFFKIPSSVPEKEKKSKCFKHLFVWCSIPSRSLQPSYHSEFI 240
Db 181 I1NNGTSIVDKGRTKTVLFFKIPSSVPEKEKKSKCFKHLFVWCSIPSRSLQPSYHSEFI 240
0Y 241 TENIVTIEOPFKIDYKLA1ATY1RGVNNASCSLSEFKKTKTWFEVDRTKKEVSTKEVT 300
Db 241 TENIVTIEOPFKIDYKLA1ATY1RGVNNASCSLSEFKKTKTWFEVDRTKKEVSTKEVT 300
0Y 301 DALVYHNHINAYEEDGHVFDIAYRDNLSYDMFYLKILDKPDEVNKKLTSPICRPFVY 360
Db 301 DALVYHNHINAYEEDGHVFDIAYRDNLSYDMFYLKILDKPDEVNKKLTSPICRPFVY 360
0Y 361 PLOYDKDAEYGSNLVLP1SAT1AVKKEKDSITCOPREILCEGIE1LPRVNYDNGKKYKYVY 420
Db 361 PLOYDKDAEYGSNLVLP1SAT1AVKKEKDSITCOPREILCEGIE1LPRVNYDNGKKYKYVY 420
0Y 421 ATEVQMSRPVPTK1AK1NVOTKEVLIHNGE0H0CWPSE1FEVSPAPREPDGAV1TCVVYSE 480
Db 421 ATEVQMSRPVPTK1AK1NVOTKEVLIHNGE0H0CWPSE1FEVSPAPREPDGAV1TCVVYSE 480
0Y 481 FNKAPF111DAKTFKE1GRATVNVEMH1DLHSM1P1NDLSAET1E 526
Db 481 FNKAPF111DAKTFKE1GRATVNVEMH1DLHSM1P1NDLSAET1E 526

```

RESULT 2

AAV97314 standard; Protein: 506 AA.

AAV97314:

01-JAN-2001 (first entry)

Beta, beta-carotene-15,15'-dioxygenase.

Beta, beta-carotene-15,15'-dioxygenase; vitamin A; beta-carotene; transformation; fruit; vegetable; developmental disorder; ophthalmological disorder; antibody; detection; quantification; treatment; therapy.

Gallus gallus.

EP1031627-A1.

30-AUG-2000.

17-FEB-2000; 2000EP-0102289.

22-FEB-1999; 99EP-0103392.

(HOFF) HOFFMANN LA KOEHE & CO AG F.

Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;

Wp1: 2000-551036/51.

Beta, beta-carotene 15,15'-dioxygenase protein, nucleic acids and antibodies, useful for production of vitamin A from carotene and gene therapy of ophthalmological disorders

Disclosure; Page 17-18; 37pp; English.

Beta, beta-carotene-15,15'-dioxygenase (bcoD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bcoD are useful as primers of probes for specific amplification and/or detection of the gene that encodes bcoD, for isolation of related sequences in other organisms, for determining bcoD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bcoD gene. The nucleotide encoding bcoD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bcoD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bcoD are used for detection/quantification of bcoD in immunoassays.

Sequence 506 AA;

Query Match 96.2%; Score 506; DR 21; Length 506;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

0Y 10 EENPEPIKAEVQGLPTWLOGVLLRNGPCMH1IGDTKYNNHWPGLALHSF1PKNKEVYVY 69
Db 1 EENPEPIKAEVQGLPTWLOGVLLRNGPCMH1IGDTKYNNHWPGLALHSF1PKNKEVYVY 69
0Y 70 PSKYT1PSHTYVNCNIEANPIVVSFEGTMAVDPCKNIFAKAFSYLSHTPEEDNGLINIM 129
Db 70 PSKYT1PSHTYVNCNIEANPIVVSFEGTMAVDPCKNIFAKAFSYLSHTPEEDNGLINIM 129
0Y 121 KTDGDDYVATSETFNTRKIDPOTLETLEKVNYSKVAAVNA1AISHRHDSACN1AISH 180
Db 121 KTDGDDYVATSETFNTRKIDPOTLETLEKVNYSKVAAVNA1AISHRHDSACN1AISH 180
0Y 190 DKGRKTVLFFKIPSSVPEKEKKSKCFKHLFVWCSIPSRSLQPSYHSEF1TENYV1TE 249
Db 190 DKGRKTVLFFKIPSSVPEKEKKSKCFKHLFVWCSIPSRSLQPSYHSEF1TENYV1TE 249
0Y 250 QPKKIDYKLA1ATY1RGVNNASCSLSEFKKTKTWFEVDRTKKEVSTKEVT1A1A1A1 309
Db 250 QPKKIDYKLA1ATY1RGVNNASCSLSEFKKTKTWFEVDRTKKEVSTKEVT1A1A1A1 309
0Y 310 NAYEEDGHVFDIAYRDNLSYDMFYLKILDKPDEVNKKLTSPICRPFVY1A1A1A1 369
Db 310 NAYEEDGHVFDIAYRDNLSYDMFYLKILDKPDEVNKKLTSPICRPFVY1A1A1A1 369

```



Query Match 3.44; Score 18; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 8 NKEEPEPIKAEEVGGQLP 25  
 |||||  
 DB 1 NKEEPEPIKAEEVGGQLP 18

RESULT 5  
 AAG62840

ID AAG62840 standard; Protein: 516 AA.

AC AAG62840;

DT 17-SEP-2001 (first entry)

DE Amino acid sequence of a beta-carotene dioxygenase (beta-diox).

KM Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;  
 transgenic plant.

US Danio rerio.

PN W0200148162-A2.

PD 05-JUL-2001.

PF 22-DEC-2000; 2030WO-E113144.

PR 24-DEC-1999; 99EP-0125895.

PR 20-MAR-2000; 2000EP-01058422.

PA (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.

PI Von Lintig J, Vogt K;

DR WP1: 2001-441713/47.

DR N-PSDB; AAH42170.

PT Dioxygenase (beta-diox) protein that cleaves beta-carotene to form  
 vitamin A aldehyde, and polynucleotides encoding them useful for  
 producing transgenic bacteria, fungi, plants expressing the polypeptide

PT producing transgenic bacteria, fungi, plants expressing the polypeptide

PS Claim 7; Page 86-87; 87pp; English.

CC The present sequence represents a beta-carotene dioxygenase (beta-diox)

CC polypeptide. Beta-diox specifically cleaves beta-carotene to form

CC vitamin A aldehyde. Beta-diox is useful for producing transgenic

CC plants. The transgenic plants have improved nutritional quality or

CC up beta-carotene condition and accumulate vitamin A aldehyde and can take

CC are useful in the study of beta-diox activity. Identification of cDNAs

CC encoding beta-diox allows the physiological characterization of

CC mammalian vitamin A metabolism. Vitamin A production in crops and

CC microorganisms can be achieved by transforming the organisms or crops

CC with the cDNAs.

CC

CC

CC

CC

CC

CC

XX ABB71747;  
 AC 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 42033.

DE Drosophila melanogaster developmental biology; cell signalling; insecticide;

DE pharmaceutical.

OS Drosophila melanogaster.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-0509231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0514150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EM;

DR WP1: 2001-655860/75.

DR N-PSDB; ABL15850.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Disclosure; SEQ ID NO 42033; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABR5737-ABR2072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pat\_sequences.

CC

CC

CC

CC

CC

CC

CC

CC

CC

RESULT 7

AAU04291

ID AAU04291 standard; Protein: 620 AA.

AC AAU04291;

DT 23-OCT-2001 (first entry)

DE Drosophila beta-carotene dioxygenase (Beta-diox 1).

KM Beta-carotene dioxygenase; beta-diox 1; Drosophila; beta-carotene;

KM lycopene; beta-apocarotenal; beta-ionone; apolycopenal; grain seed; corn;

KM oil seed; palm; chick-pea; diagnostic; therapeutic; ribosome;

KM retinoid/vitamin A deficiency; beta-diox 1; Transgenic; nutrition;

KM carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop;

XX

OS Drosophila melanogaster.  
 XX  
 PN W0200148163-A2.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 27-DEC-2000; 2000MC-EP13273.  
 XX  
 PR 24-DEC-1999; 99EP-0125895.  
 FF 20-MAR-2000; 2000EP-0105822.  
 XX  
 PA (GREE-) GPFENOVATION PFLANZENBIOTECHEMLOGIE GMBH  
 PI Von Linde J, Vogt K;  
 XX  
 DR WPI: 2001-425657/45.  
 XX  
 N-PSDB: AAS07180.  
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase  
 PT (beta-diox II) protein that cleaves beta-carotene and lycopene to yield  
 PT beta-apocarothenal and beta-ionone, and apolycopeneals, respectively  
 XX  
 PS Disclosure: Fig 6; 116pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of Drosophila beta-  
 CC carotene dioxygenase (Beta-diox II). Beta-diox specifically cleaves beta-  
 CC carotene and lycopene to form beta-apocarothenal and beta-ionone, and  
 CC apolycopeneals, respectively. The DNA is useful for transforming grain  
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;  
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes;  
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or  
 CC ribozyme type therapeutic agents and for detecting any abnormality of  
 CC endogenous beta-diox II. The beta-diox II specific antisense  
 CC oligonucleotides derived from the DNA sequence are useful for dose  
 CC response studies in relevant models of retinoid/vitamin A deficiency  
 CC during any stage of an organism's development. The nucleic acids are also  
 CC useful as probes and as a guideline to define new PCR (polymerase chain  
 CC reaction) primers for the cloning of substantially homologous DNA  
 CC sequences from other sources. The nucleic acids are also useful for  
 CC determining the presence or quantity of beta-diox II nucleic acid and  
 CC determining presence and amount of beta-diox II. The polypeptide is also  
 CC useful for increasing or decreasing the level of vitamin A and other  
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,  
 CC screening of an expression library to identify nucleic acids encoding  
 CC beta-diox II or the structure of functional domains. The transgene  
 CC lungt, yeast, insect, animal or plant cells, seeds, tissues, or whole  
 CC organisms have improved nutritional quality or physiological condition  
 CC and accumulate important metabolites of carotene/retinoid pathways such  
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta  
 CC carotene from the medium. Expression systems encoding beta-diox II are  
 CC useful in the study of beta-diox II activity. Identification of cDNAs  
 CC encoding beta-diox I and II allows the physiological characterisation of  
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 SQ Sequence 620 AA;  
 Query Match 2.1%; Score 11; DB 22; Length 620;  
 Best Local Similarity 100.0%; Pred No. 0.047;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 453 PSEPIVPSPD 463  
 IIIIIIIIIIII  
 DB 550 PSEPIVPSPD 560

AC AAG62841;  
 XX  
 DT 17-SEP 2001 (first entry)  
 XX  
 DE Amino acid sequence of a beta-carotene dioxygenase (beta diox).  
 XX  
 KW Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;  
 KW transgenic plant.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200148162-A2.  
 XX  
 PD 05-JUL-2001  
 XX  
 PR 24-DEC-2000; 2000MC-EP13144.  
 XX  
 FF 20-MAR-2000; 2000EP-0105822.  
 XX  
 PA (GREE-) GPFENOVATION PFLANZENBIOTECHEMLOGIE GMBH.  
 PI Von Linde J, Vogt K;  
 XX  
 DR WPI: 2001-441713/47.  
 XX  
 N-PSDB: AAG42156.  
 PT D-oxyg-nase (beta-diox) protein that cleaves beta-carotene to form  
 PT vitamin A aldehyde, and polynucleotides encoding them useful for  
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide  
 XX  
 PS Claim 7; Fig 6; 87pp; English.  
 XX  
 CC The present sequence represents a beta-carotene dioxygenase (beta-diox)  
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form  
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic  
 CC plants. The transgenic plants have improved nutritional quality or  
 CC physiological condition and accumulate vitamin A aldehyde and can take  
 CC up beta-carotene from the medium. Expression systems encoding beta-diox  
 CC are useful in the study of beta-diox activity. Identification of cDNAs  
 CC encoding beta-diox allows the physiological characterisation of  
 CC mammalian vitamin A metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 SQ Sequence 620 AA;  
 Query Match 2.1%; Score 11; DB 22; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 0.047;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 453 PSEPIVPSPD 463  
 IIIIIIIIIIII  
 DB 550 PSEPIVPSPD 560

## RESULT 9

AAU04293

ID AAU04293 standard; Protein; 549 AA.

AC AAU04293;

XX 23-OCT-2001 (first entry)

DE Zebra-2, beta-carotene dioxygenase (beta-diox II).

KW Beta-carotene dioxygenase, beta-diox I, zebra-2, beta-carotene;

KW lycopene; beta-apocarothenal; beta-ionone; apolycopeneal; grain seed; corn;

KW oil seed; palm; chick-pea; diagnostic; therapeutic; ribozyme; zebrafish;

KW retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;

KW carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.

XX

08 Brachydanio rerio.  
 XX  
 PN W020014816-A2.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 27-DEC-2000: 2000W0-EPI3273  
 XX  
 PR 24 DEC 1999: 99EP-0125895.  
 PR 20 MAR 2000: 2000EP-0105822  
 XX  
 PA (GREF.) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH  
 XX  
 PI Von Lintig J, Vogt K;  
 XX  
 DR WPI: 2001-425657/45.  
 DR N-PSDB: AAS07195.  
 XX  
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase  
 PT (beta-diox II) protein that cleaves beta-carotene and lycopene to yield  
 PT beta-apocrotene and beta-ionone, and apolycopene, respectively  
 PS  
 PS Claim 2: Fig 14: 116pp: English.  
 CC The sequence represents the amino acid sequence of zebra-2, beta-carotene  
 CC dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene  
 CC and lycopene to form beta-apocrotene and beta-ionone, and  
 CC apolycopene, respectively. The DNA is useful for transforming grain  
 CC crops, e.g. corn; oil seeds, e.g. Brassica seeds, edible seeds e.g. palm;  
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes,  
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or  
 CC ribozyme type therapeutic agents and for detecting any abnormality of  
 CC endogenous beta-diox II. The beta-diox II specific antisense  
 CC oligonucleotides derived from the DNA sequence are useful for dose  
 CC response studies in relevant models of retinoid/vitamin A deficiency  
 CC during any stage of an organism's development. The nucleic acids are also  
 CC useful as probes and as a guideline to define new PCR (polymerase chain  
 CC reaction) primers for the cloning of substantially homologous DNA  
 CC sequences from other sources. The nucleic acids are also useful for  
 CC determining the presence or quantity of beta-diox II nucleic acid and  
 CC determining presence and amount of beta-diox II. The polypeptide is also  
 CC useful for increasing or decreasing the amount of beta-diox II levels in  
 CC a cell or tissue which can modify the level of vitamin A and other  
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,  
 CC screening of an expression library to identify nucleic acids encoding  
 CC beta-diox II or the structure of functional domains. The transgenic  
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole  
 CC organisms have improved nutritional quality or physiological condition  
 CC and accumulate important metabolites of carotene/retinoid pathways such  
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta-  
 CC carotene from the medium. Expression systems encoding beta-diox II are  
 CC useful in the study of beta-diox II activity. Identification of cDNAs  
 CC encoding beta-diox I and II allows the physiological characterisation of  
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 XX Sequence 549 AA:  
 SO  
 Query Match 1.98: Score 10: DR 22: Length 549:  
 Best Local Similarity 100.0%: Prod No 0.43:  
 Matches 10: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 OY 86 NRIVSSEFCT 95  
 DB 111 NRIVSSEFCT 120  
 RESULT 10  
 ID AAY97312 standard; Peptide: 8 AA.  
 XX

AC AAY97312:  
 XX  
 PT 03 JAN 2001 (first entry)  
 XX  
 DE Beta, Beta-carotene-15,15-dioxygenase tryptic peptide treatment.  
 XX  
 KW Beta, beta-carotene 15,15-dioxygenase; vitamin A; beta-carotene;  
 KW transformation; fruit; vegetable; developmental disorder;  
 KW ophthalmological disorders; antibody; detection; quantitation;  
 KW treatment; therapy.  
 XX  
 OS Gallus gallus.  
 XX  
 PN EP1031627-A1.  
 XX  
 PD 30-AUG-2000.  
 XX  
 PF 17-FEB-2000: 2000EP-0102289.  
 XX  
 PR 22-FEB-1999: 99EP-0103482.  
 XX  
 PA (HOFF.) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;  
 PI Wyss M;  
 XX  
 DR WPI: 2000-551036/51.  
 XX  
 PT Beta-beta-carotene 15,15-dioxygenase protein, nucleic acids and  
 PT antibodies, useful for production of vitamin A from carotene and gene  
 PT therapy of ophthalmological disorders  
 PS  
 PS Example 3, Page 7: 37pp: English.  
 CC Beta, beta-carotene-15,15-dioxygenase (bctd) is used for enzymatic  
 CC conversion of beta-carotene to vitamin A and for raising specific  
 CC antibodies. Nucleotides encoding all or part of bctd are useful as  
 CC primers or probes for specific amplification and/or detection of  
 CC the gene that encodes bctd, for isolation of related sequences in  
 CC other organisms, for determining bctd levels in humans (to identify  
 CC subjects requiring vitamin A supplementation) and for detecting  
 CC mutations in the bctd gene. The nucleotide encoding bctd may also  
 CC be used to transform cells, particularly plant cells, to increase  
 CC their vitamin A contents (especially in fruits and vegetables) and  
 CC in gene therapy of subjects who have mutated or deleted forms of the  
 CC bctd gene and thus low vitamin A levels and susceptibility to  
 CC developmental or ophthalmological disorders. Antibodies raised  
 CC against bctd are used for detection/quantification of bctd in  
 CC immunoassays.  
 XX  
 XX Sequence 8 AA:  
 SO  
 Query Match 1.58: Score 8: DR 21: Length 8:  
 Best Local Similarity 100.0%: Prod No 7.6005:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 OY 18 AEVGGQLP 25  
 DB 1 AEVGGQLP 8  
 RESULT 11  
 ID ABG28616 standard; Protein: 46 AA.  
 XX  
 AC ABG28616:  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #28607.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnosis; genetic disorder.  
 KW

XX OS Homo sapiens.  
 XX PN W0200175067-A2  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US06631.  
 XX PR 31-MAR-2000; 2000US-0540217  
 XX PR 23-AUG-2000; 2000US-0649167  
 XX PA (HVSF-) HVSFO INC.  
 XX PI Drmanac RT, Liu C, Tang YF;  
 XX DR WPI: 2601-63462/73.  
 XX DR N-PSDB; AAS92803.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PS Claim 20; SEQ ID No 58975; 103pp; English.  
 XX XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antipeptides against (I), detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAG50010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIP0  
 CC at ftp.wipo.int/pub/published\_pt\_sequences.  
 XX XX  
 SO Sequence 46 AA;  
 QY 223 SIPSRSL 230  
 DB 25 SIPSRSL 32  
 11111111  
 1.5%; Score 8; DB 22; Length 46;  
 Query Match Best Local Similarity 100.0%; Pred No 4 4;  
 Matches 8; Conservation 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 12  
 AAY60294  
 ID AAY60294 standard; Protein; 82 AA.  
 XX  
 AC AAY60294;  
 XX  
 DT 31-JAN-2000 (first entry)  
 XX  
 DE Human endometrium tumour EST encoded protein 354.  
 XX  
 DE Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
 KW treatment; uterine; gene therapy; expressed sequence tag  
 XX  
 OS Homo sapiens

XX PN DE19817948-A1.  
 XX PD 21-OCT-1999  
 XX PR 17-APR-1998; 98DE-1017948.  
 XX PR 17-APR-1998; 98DE-1017948.  
 XX PA (MEIA-) METAGEN GEN GENOMEFORSCHUNG MBH.  
 XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl H;  
 XX DR WPI: 1999-591957/51.  
 XX DR N-PSDB; AA242095.  
 PT New nucleic acid sequences expressed in uterine cancer tissues, and  
 PT derived polypeptides, for treatment of uterine and endometrial cancer  
 PT and identification of therapeutic agents.  
 PS Claim 23; Page 416; 444pp; German.  
 XX XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for  
 CC generation of specific antibodies. (A) are identified by assembling ESTs  
 CC (expressed sequence tags) from a particular tissue type before comparison  
 CC of expression patterns. This allows a significantly longer fragment of  
 CC the gene to be revealed, so should reduce the number of failures  
 CC associated with the 192 base ESTs from different libraries way represent  
 CC different parts of the same unknown gene, distorting the estimated  
 CC frequency of occurrence in a particular tissue. AAY5941 Y60328 represent  
 CC protein fragments encoded by the human endometrium tumour cDNA library  
 CC derived EST fragments represented in AAC41981-242121.  
 XX XX  
 SO Sequence 82 AA;  
 QY 21 QGQLPTWL 28  
 DB 20 QGQLPTWL 27  
 11111111  
 1.5%; Score 8; DB 20; Length 82;  
 Query Match Best Local Similarity 100.0%; Pred No. 7 6;  
 Matches 8; Conservation 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 13  
 AAR73956  
 ID AAR73956 standard; Protein; 205 AA.  
 XX  
 AC AAR73956;  
 XX  
 DT 29 MAY-2001 (first entry)  
 XX  
 DE Canine mutant RPE65.  
 XX  
 DE Canine, dog, RPE65, canine retinal pigment epithelium 65, RPE65;  
 KW congenital stationary night blindness; CSNB; mutation detection;  
 KW mutant.  
 XX  
 OS Canis familiaris.  
 XX  
 PN US6201114-B1.  
 XX  
 PD 13-MAR-2001.  
 XX  
 DE 36 Aug 1999; 98US 0686259.  
 XX  
 OS 06-OCT-1999; 98US-0103219.

XX (CORK) CORNELL RES FOUND INC.  
 PA Aguirre GD, Acland GM, Ray K;  
 XX WPI: 2001-265168/27.  
 XX Novel nucleic acid molecule encoding canine retinal pigment epithelium  
 PT 65, where presence of mutation in one or both alleles is indicative of  
 PT a carrier of, or dog affected with congenital stationary night  
 PT blindness.  
 XX Disclosure: Fig 1C; 25pp; English.  
 XX The present sequence is a mutated version of canine retinal pigment  
 CC epithelium 65 (PPE65). The gene encoding this protein contains an AACA  
 CC deletion at nucleotides 487-490 and is responsible for the disease  
 CC congenital stationary night blindness (CSNB). The gene is useful for  
 CC identifying dogs which are genetically normal, or are carriers of, or  
 CC affected with congenital stationary night blindness (CSNB). This allows  
 CC a breeder to eliminate the carrier from the breeding stock or to breed  
 CC carriers with genetically normal dogs.  
 XX Sequence: 205 AA;  
 SQ  
 Query Match 1.58, Score 8, DB 22, Length: 205;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 99 PDPCKNIF 106  
 IDB 109 PDPCKNIF 116  
 RESULT 14  
 AAG41938  
 ID AAG41938 standard; Protein: 318 AA.  
 XX AAG41938;  
 AC  
 XX 18-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 52240.  
 XX  
 XX Protein identification: signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 XX Arabidopsis thaliana.  
 OS  
 XX EPI034405-A2.  
 EN  
 XX 06-SEP-2000.  
 PD  
 XX 25-FEB-2000, 2000PP 0101419  
 XX  
 XX 25-FEB-1999; 990S-0121825;  
 PR 05-MAR-1999; 990S-0123180;  
 PR 09-MAR-1999; 990S-0123548;  
 PR 23-MAR-1999; 990S-0125788;  
 PR 25-MAR-1999; 990S-0126264;  
 PR 29-MAR-1999; 990S-0126785;  
 PR 01-APR-1999; 990S-0127462;  
 PR 06-APR-1999; 990S-0128234;  
 PR 08-APR-1999; 990S-0128714;  
 PR 16-APR-1999; 990S-0129845;  
 PR 19-APR-1999; 990S-0130077;  
 PR 21-APR-1999; 990S-0130449;  
 PR 23-APR-1999; 990S-0130510;  
 PR 28-APR-1999; 990S-0130891;  
 PR 30-APR-1999; 990S-0131449;  
 PR 30-APR-1999; 990S-0132048;  
 PR 30-APR-1999; 990S-0132407.

PR 04-MAY-1999; 990S-0132484;  
 PR 05-MAY-1999; 990S-0132485;  
 PR 06-MAY-1999; 990S-0132486;  
 PR 06-MAY-1999; 990S-0132487;  
 PR 07-MAY-1999; 990S-0132863;  
 PR 11-MAY-1999; 990S-0134256;  
 PR 14-MAY-1999; 990S-0134218;  
 PR 14-MAY-1999; 990S-0134219;  
 PR 14-MAY-1999; 990S-0134221;  
 PR 14-MAY-1999; 990S-0134370;  
 PR 18-MAY-1999; 990S-0134768;  
 PR 19-MAY-1999; 990S-0134941;  
 PR 20-MAY-1999; 990S-0135124;  
 PR 21-MAY-1999; 990S-0135353;  
 PR 24-MAY-1999; 990S-0135629;  
 PR 25-MAY-1999; 990S-0136021;  
 PR 27-MAY-1999; 990S-0136392;  
 PR 28-MAY-1999; 990S-0136782;  
 PR 01-JUN-1999; 990S-0137222;  
 PR 01-JUN-1999; 990S-0137528;  
 PR 03-JUN-1999; 990S-0137502;  
 PR 07-JUN-1999; 990S-0137724;  
 PR 08-JUN-1999; 990S-0138064;  
 PR 10-JUN-1999; 990S-0138540;  
 PR 10-JUN-1999; 990S-0138847;  
 PR 14-JUN-1999; 990S-0139119;  
 PR 16-JUN-1999; 990S-0139452;  
 PR 16-JUN-1999; 990S-0139453;  
 PR 17-JUN-1999; 990S-0139492;  
 PR 18-JUN-1999; 990S-0139454;  
 PR 18-JUN-1999; 990S-0139455;  
 PR 18-JUN-1999; 990S-0139456;  
 PR 18-JUN-1999; 990S-0139457;  
 PR 18-JUN-1999; 990S-0139458;  
 PR 18-JUN-1999; 990S-0139459;  
 PR 18-JUN-1999; 990S-0139460;  
 PR 18-JUN-1999; 990S-0139461;  
 PR 18-JUN-1999; 990S-0139462;  
 PR 18-JUN-1999; 990S-0139463;  
 PR 18-JUN-1999; 990S-0139464;  
 PR 18-JUN-1999; 990S-0139465;  
 PR 18-JUN-1999; 990S-0139466;  
 PR 18-JUN-1999; 990S-0139467;  
 PR 18-JUN-1999; 990S-0139468;  
 PR 18-JUN-1999; 990S-0139469;  
 PR 18-JUN-1999; 990S-0139470;  
 PR 18-JUN-1999; 990S-0139471;  
 PR 18-JUN-1999; 990S-0139472;  
 PR 18-JUN-1999; 990S-0139473;  
 PR 18-JUN-1999; 990S-0139474;  
 PR 18-JUN-1999; 990S-0139475;  
 PR 18-JUN-1999; 990S-0139476;  
 PR 18-JUN-1999; 990S-0139477;  
 PR 18-JUN-1999; 990S-0139478;  
 PR 18-JUN-1999; 990S-0139479;  
 PR 18-JUN-1999; 990S-0139480;  
 PR 18-JUN-1999; 990S-0139481;  
 PR 18-JUN-1999; 990S-0139482;  
 PR 18-JUN-1999; 990S-0139483;  
 PR 18-JUN-1999; 990S-0139484;  
 PR 18-JUN-1999; 990S-0139485;  
 PR 18-JUN-1999; 990S-0139486;  
 PR 18-JUN-1999; 990S-0139487;  
 PR 18-JUN-1999; 990S-0139488;  
 PR 18-JUN-1999; 990S-0139489;  
 PR 18-JUN-1999; 990S-0139490;  
 PR 18-JUN-1999; 990S-0139491;  
 PR 18-JUN-1999; 990S-0139492;  
 PR 18-JUN-1999; 990S-0139493;  
 PR 18-JUN-1999; 990S-0139494;  
 PR 18-JUN-1999; 990S-0139495;  
 PR 18-JUN-1999; 990S-0139496;  
 PR 18-JUN-1999; 990S-0139497;  
 PR 18-JUN-1999; 990S-0139498;  
 PR 18-JUN-1999; 990S-0139499;  
 PR 18-JUN-1999; 990S-0139500;  
 PR 18-JUN-1999; 990S-0139501;  
 PR 18-JUN-1999; 990S-0139502;  
 PR 18-JUN-1999; 990S-0139503;  
 PR 18-JUN-1999; 990S-0139504;  
 PR 18-JUN-1999; 990S-0139505;  
 PR 18-JUN-1999; 990S-0139506;  
 PR 18-JUN-1999; 990S-0139507;  
 PR 18-JUN-1999; 990S-0139508;  
 PR 18-JUN-1999; 990S-0139509;  
 PR 18-JUN-1999; 990S-0139510;  
 PR 18-JUN-1999; 990S-0139511;  
 PR 18-JUN-1999; 990S-0139512;  
 PR 18-JUN-1999; 990S-0139513;  
 PR 18-JUN-1999; 990S-0139514;  
 PR 18-JUN-1999; 990S-0139515;  
 PR 18-JUN-1999; 990S-0139516;  
 PR 18-JUN-1999; 990S-0139517;  
 PR 18-JUN-1999; 990S-0139518;  
 PR 18-JUN-1999; 990S-0139519;  
 PR 18-JUN-1999; 990S-0139520;  
 PR 18-JUN-1999; 990S-0139521;  
 PR 18-JUN-1999; 990S-0139522;  
 PR 18-JUN-1999; 990S-0139523;  
 PR 18-JUN-1999; 990S-0139524;  
 PR 18-JUN-1999; 990S-0139525;  
 PR 18-JUN-1999; 990S-0139526;  
 PR 18-JUN-1999; 990S-0139527;  
 PR 18-JUN-1999; 990S-0139528;  
 PR 18-JUN-1999; 990S-0139529;  
 PR 18-JUN-1999; 990S-0139530;  
 PR 18-JUN-1999; 990S-0139531;  
 PR 18-JUN-1999; 990S-0139532;  
 PR 18-JUN-1999; 990S-0139533;  
 PR 18-JUN-1999; 990S-0139534;  
 PR 18-JUN-1999; 990S-0139535;  
 PR 18-JUN-1999; 990S-0139536;  
 PR 18-JUN-1999; 990S-0139537;  
 PR 18-JUN-1999; 990S-0139538;  
 PR 18-JUN-1999; 990S-0139539;  
 PR 18-JUN-1999; 990S-0139540;  
 PR 18-JUN-1999; 990S-0139541;  
 PR 18-JUN-1999; 990S-0139542;  
 PR 18-JUN-1999; 990S-0139543;  
 PR 18-JUN-1999; 990S-0139544;  
 PR 18-JUN-1999; 990S-0139545;  
 PR 18-JUN-1999; 990S-0139546;  
 PR 18-JUN-1999; 990S-0139547;  
 PR 18-JUN-1999; 990S-0139548;  
 PR 18-JUN-1999; 990S-0139549;  
 PR 18-JUN-1999; 990S-0139550;  
 PR 18-JUN-1999; 990S-0139551;  
 PR 18-JUN-1999; 990S-0139552;  
 PR 18-JUN-1999; 990S-0139553;  
 PR 18-JUN-1999; 990S-0139554;  
 PR 18-JUN-1999; 990S-0139555;  
 PR 18-JUN-1999; 990S-0139556;  
 PR 18-JUN-1999; 990S-0139557;  
 PR 18-JUN-1999; 990S-0139558;  
 PR 18-JUN-1999; 990S-0139559;  
 PR 18-JUN-1999; 990S-0139560;  
 PR 18-JUN-1999; 990S-0139561;  
 PR 18-JUN-1999; 990S-0139562;  
 PR 18-JUN-1999; 990S-0139563;  
 PR 18-JUN-1999; 990S-0139564;  
 PR 18-JUN-1999; 990S-0139565;  
 PR 18-JUN-1999; 990S-0139566;  
 PR 18-JUN-1999; 990S-0139567;  
 PR 18-JUN-1999; 990S-0139568;  
 PR 18-JUN-1999; 990S-0139569;  
 PR 18-JUN-1999; 990S-0139570;  
 PR 18-JUN-1999; 990S-0139571;  
 PR 18-JUN-1999; 990S-0139572;  
 PR 18-JUN-1999; 990S-0139573;  
 PR 18-JUN-1999; 990S-0139574;  
 PR 18-JUN-1999; 990S-0139575;  
 PR 18-JUN-1999; 990S-0139576;  
 PR 18-JUN-1999; 990S-0139577;  
 PR 18-JUN-1999; 990S-0139578;  
 PR 18-JUN-1999; 990S-0139579;  
 PR 18-JUN-1999; 990S-0139580;  
 PR 18-JUN-1999; 990S-0139581;  
 PR 18-JUN-1999; 990S-0139582;  
 PR 18-JUN-1999; 990S-0139583;  
 PR 18-JUN-1999; 990S-0139584;  
 PR 18-JUN-1999; 990S-0139585;  
 PR 18-JUN-1999; 990S-0139586;  
 PR 18-JUN-1999; 990S-0139587;  
 PR 18-JUN-1999; 990S-0139588;  
 PR 18-JUN-1999; 990S-0139589;  
 PR 18-JUN-1999; 990S-0139590;  
 PR 18-JUN-1999; 990S-0139591;  
 PR 18-JUN-1999; 990S-0139592;  
 PR 18-JUN-1999; 990S-0139593;  
 PR 18-JUN-1999; 990S-0139594;  
 PR 18-JUN-1999; 990S-0139595;  
 PR 18-JUN-1999; 990S-0139596;  
 PR 18-JUN-1999; 990S-0139597;  
 PR 18-JUN-1999; 990S-0139598;  
 PR 18-JUN-1999; 990S-0139599;  
 PR 18-JUN-1999; 990S-0139600;  
 PR 18-JUN-1999; 990S-0139601;  
 PR 18-JUN-1999; 990S-0139602;  
 PR 18-JUN-1999; 990S-0139603;  
 PR 18-JUN-1999; 990S-0139604;  
 PR 18-JUN-1999; 990S-0139605;  
 PR 18-JUN-1999; 990S-0139606;  
 PR 18-JUN-1999; 990S-0139607;  
 PR 18-JUN-1999; 990S-0139608;  
 PR 18-JUN-1999; 990S-0139609;  
 PR 18-JUN-1999; 990S-0139610;  
 PR 18-JUN-1999; 990S-0139611;  
 PR 18-JUN-1999; 990S-0139612;  
 PR 18-JUN-1999; 990S-0139613;  
 PR 18-JUN-1999; 990S-0139614;  
 PR 18-JUN-1999; 990S-0139615;  
 PR 18-JUN-1999; 990S-0139616;  
 PR 18-JUN-1999; 990S-0139617;  
 PR 18-JUN-1999; 990S-0139618;  
 PR 18-JUN-1999; 990S-0139619;  
 PR 18-JUN-1999; 990S-0139620;  
 PR 18-JUN-1999; 990S-0139621;  
 PR 18-JUN-1999; 990S-0139622;  
 PR 18-JUN-1999; 990S-0139623;  
 PR 18-JUN-1999; 990S-0139624;  
 PR 18-JUN-1999; 990S-0139625;  
 PR 18-JUN-1999; 990S-0139626;  
 PR 18-JUN-1999; 990S-0139627;  
 PR 18-JUN-1999; 990S-0139628;  
 PR 18-JUN-1999; 990S-0139629;  
 PR 18-JUN-1999; 990S-0139630;  
 PR 18-JUN-1999; 990S-0139631;  
 PR 18-JUN-1999; 990S-0139632;  
 PR 18-JUN-1999; 990S-0139633;  
 PR 18-JUN-1999; 990S-0139634;  
 PR 18-JUN-1999; 990S-0139635;  
 PR 18-JUN-1999; 990S-0139636;  
 PR 18-JUN-1999; 990S-0139637;  
 PR 18-JUN-1999; 990S-0139638;  
 PR 18-JUN-1999; 990S-0139639;  
 PR 18-JUN-1999; 990S-0139640;  
 PR 18-JUN-1999; 990S-0139641;  
 PR 18-JUN-1999; 990S-0139642;  
 PR 18-JUN-1999; 990S-0139643;  
 PR 18-JUN-1999; 990S-0139644;  
 PR 18-JUN-1999; 990S-0139645;  
 PR 18-JUN-1999; 990S-0139646;  
 PR 18-JUN-1999; 990S-0139647;  
 PR 18-JUN-1999; 990S-0139648;  
 PR 18-JUN-1999; 990S-0139649;  
 PR 18-JUN-1999; 990S-0139650;  
 PR 18-JUN-1999; 990S-0139651;  
 PR 18-JUN-1999; 990S-0139652;  
 PR 18-JUN-1999; 990S-0139653;  
 PR 18-JUN-1999; 990S-0139654;  
 PR 18-JUN-1999; 990S-0139655;  
 PR 18-JUN-1999; 990S-0139656;  
 PR 18-JUN-1999; 990S-0139657;  
 PR 18-JUN-1999; 990S-0139658;  
 PR 18-JUN-1999; 990S-0139659;  
 PR 18-JUN-1999; 990S-0139660;  
 PR 18-JUN-1999; 990S-0139661;  
 PR 18-JUN-1999; 990S-0139662;  
 PR 18-JUN-1999; 990S-0139663;  
 PR 18-JUN-1999; 990S-0139664;  
 PR 18-JUN-1999; 990S-0139665;  
 PR 18-JUN-1999; 990S-0139666;  
 PR 18-JUN-1999; 990S-0139667;  
 PR 18-JUN-1999; 990S-0139668;  
 PR 18-JUN-1999; 990S-0139669;  
 PR 18-JUN-1999; 990S-0139670;  
 PR 18-JUN-1999; 990S-0139671;  
 PR 18-JUN-1999; 990S-0139672;  
 PR 18-JUN-1999; 990S-0139673;  
 PR 18-JUN-1999; 990S-0139674;  
 PR 18-JUN-1999; 990S-0139675;  
 PR 18-JUN-1999; 990S-0139676;  
 PR 18-JUN-1999; 990S-0139677;  
 PR 18-JUN-1999; 990S-0139678;  
 PR 18-JUN-1999; 990S-0139679;  
 PR 18-JUN-1999; 990S-0139680;  
 PR 18-JUN-1999; 990S-0139681;  
 PR 18-JUN-1999; 990S-0139682;  
 PR 18-JUN-1999; 990S-0139683;  
 PR 18-JUN-1999; 990S-0139684;  
 PR 18-JUN-1999; 990S-0139685;  
 PR 18-JUN-1999; 990S-0139686;  
 PR 18-JUN-1999; 990S-0139687;  
 PR 18-JUN-1999; 990S-0139688;  
 PR 18-JUN-1999; 990S-0139689;  
 PR 18-JUN-1999; 990S-0139690;  
 PR 18-JUN-1999; 990S-0139691;  
 PR 18-JUN-1999; 990S-0139692;  
 PR 18-JUN-1999; 990S-0139693;  
 PR 18-JUN-1999; 990S-0139694;  
 PR 18-JUN-1999; 990S-0139695;  
 PR 18-JUN-1999; 990S-0139696;  
 PR 18-JUN-1999; 990S-0139697;  
 PR 18-JUN-1999; 990S-0139698;  
 PR 18-JUN-1999; 990S-0139699;  
 PR 18-JUN-1999; 990S-0139700;  
 PR 18-JUN-1999; 990S-0139701;  
 PR 18-JUN-1999; 990S-0139702;  
 PR 18-JUN-1999; 990S-0139703;  
 PR 18-JUN-1999; 990S-0139704;  
 PR 18-JUN-1999; 990S-0139705;  
 PR 18-JUN-1999; 990S-0139706;  
 PR 18-JUN-1999; 990S-0139707;  
 PR 18-JUN-1999; 990S-0139708;  
 PR 18-JUN-1999; 990S-0139709;  
 PR 18-JUN-1999; 990S-0139710;  
 PR 18-JUN-1999; 990S-0139711;  
 PR 18-JUN-1999; 990S-0139712;  
 PR 18-JUN-1999; 990S-0139713;  
 PR 18-JUN-1999; 990S-0139714;  
 PR 18-JUN-1999; 990S-0139715;  
 PR 18-JUN-1999; 990S-0139716;  
 PR 18-JUN-1999; 990S-0139717;  
 PR 18-JUN-1999; 990S-0139718;  
 PR 18-JUN-1999; 990S-0139719;  
 PR 18-JUN-1999; 990S-0139720;  
 PR 18-JUN-1999; 990S-0139721;  
 PR 18-JUN-1999; 990S-0139722;  
 PR 18-JUN-1999; 990S-0139723;  
 PR 18-JUN-1999; 990S-0139724;  
 PR 18-JUN-1999; 990S-0139725;  
 PR 18-JUN-1999; 990S-0139726;  
 PR 18-JUN-1999; 990S-0139727;  
 PR 18-JUN-1999; 990S-0139728;  
 PR 18-JUN-1999; 990S-0139729;  
 PR 18-JUN-1999; 990S-0139730;  
 PR 18-JUN-1999; 990S-0139731;  
 PR 18-JUN-1999; 990S-0139732;  
 PR 18-JUN-1999; 990S-0139733;  
 PR 18-JUN-1999; 990S-0139734;  
 PR 18-JUN-1999; 990S-0139735;  
 PR 18-JUN-1999; 990S-0139736;  
 PR 18-JUN-1999; 990S-0139737;  
 PR 18-JUN-1999; 990S-0139738;  
 PR 18-JUN-1999; 990S-0139739;  
 PR 18-JUN-1999; 990S-0139740;  
 PR 18-JUN-1999; 990S-0139741;  
 PR 18-JUN-1999; 990S-0139742;  
 PR 18-JUN-1999; 990S-0139743;  
 PR 18-JUN-1999; 990S-0139744;  
 PR 18-JUN-1999; 990S-0139745;  
 PR 18-JUN-1999; 990S-0139746;  
 PR 18-JUN-1999; 990S-0139747;  
 PR 18-JUN-1999; 990S-0139748;  
 PR 18-JUN-1999; 990S-0139749;  
 PR 18-JUN-1999; 990S-0139750;  
 PR 18-JUN-1999; 990S-0139751;  
 PR 18-JUN-1999; 990S-0139752;  
 PR 18-JUN-1999; 990S-0139753;  
 PR 18-JUN-1999; 990S-0139754;  
 PR 18-JUN-1999; 990S-0139755;  
 PR 18-JUN-1999; 990S-0139756;  
 PR 18-JUN-1999; 990S-0139757;  
 PR 18-JUN-1999; 990S-0139758;  
 PR 18-JUN-1999; 990S-0139759;  
 PR 18-JUN-1999; 990S-0139760;  
 PR 18-JUN-1999; 990S-0139761;  
 PR 18-JUN-1999; 990S-0139762;  
 PR 18-JUN-1999; 990S-0139763;  
 PR 18-JUN-1999; 990S-0139764;  
 PR 18-JUN-1999; 990S-0139765;  
 PR 18-JUN-1999; 990S-0139766;  
 PR 18-JUN-1999; 990S-0139767;  
 PR 18-JUN-1999; 990S-0139768;  
 PR 18-JUN-1999; 990S-0139769;  
 PR 18-JUN-1999; 990S-0139770;  
 PR 18-JUN-1999; 990S-0139771;  
 PR 18-JUN-1999; 990S-0139772;  
 PR 18-JUN-1999; 990S-0139773;  
 PR 18-JUN-1999; 990S-0139774;  
 PR 18-JUN-1999; 990S-0139775;  
 PR 18-JUN-1999; 990S-0139776;  
 PR 18-JUN-1999; 990S-0139777;  
 PR 18-JUN-1999; 990S-0139778;  
 PR 18-JUN-1999; 990S-0139779;  
 PR 18-JUN-1999; 990S-0139780;  
 PR 18-JUN-1999; 990S-0139781;  
 PR 18-JUN-1999; 990S-0139782;  
 PR 18-JUN-1999; 990S-0139783;  
 PR 18-JUN-1999; 990S-0139784;  
 PR 18-JUN-1999; 990S-0139785;  
 PR 18-JUN-1999; 990S-0139786;  
 PR 18-JUN-1999; 990S-0139787;  
 PR 18-JUN-1999; 990S-0139788;  
 PR 18-JUN-1999; 990S-0139789;  
 PR 18-JUN-1999; 990S-0139790;  
 PR 18-JUN-1999; 990S-01397



PR 22-JUL-1999: 99US-0145085  
 PR 22-JUL-1999: 99US-0145087  
 PR 22-JUL-1999: 99US-0145089  
 PR 22-JUL-1999: 99US-0145192  
 PR 23-JUL-1999: 99US-0145145  
 PR 23-JUL-1999: 99US-0145218  
 PR 23-JUL-1999: 99US-0145224  
 PR 26-JUL-1999: 99US-0145276  
 PR 27-JUL-1999: 99US-0145913  
 PR 27-JUL-1999: 99US-0145918  
 PR 28-JUL-1999: 99US-0145951  
 PR 02-AUG-1999: 99US-0146386  
 PR 02-AUG-1999: 99US-0146389  
 PR 03-AUG-1999: 99US-0147034  
 PR 04-AUG-1999: 99US-0147204  
 PR 05-AUG-1999: 99US-0147302  
 PR 05-AUG-1999: 99US-0147392  
 PR 06-AUG-1999: 99US-0147393  
 PR 06-AUG-1999: 99US-0147416  
 PR 09-AUG-1999: 99US-0147493  
 PR 09-AUG-1999: 99US-0147935  
 PR 10-AUG-1999: 99US-0148171  
 PR 11-AUG-1999: 99US-0148319  
 PR 12-AUG-1999: 99US-0148341  
 PR 13-AUG-1999: 99US-0148565  
 PR 13-AUG-1999: 99US-0148684  
 PR 16-AUG-1999: 99US-0149468  
 PR 17-AUG-1999: 99US-0149475  
 PR 18-AUG-1999: 99US-0149426  
 PR 20-AUG-1999: 99US-0149722  
 PR 20-AUG-1999: 99US-0149723  
 PR 20-AUG-1999: 99US-0149929  
 PR 23-AUG-1999: 99US-0149902  
 PR 23-AUG-1999: 99US-0149930  
 PR 25-AUG-1999: 99US-0150566  
 PR 26-AUG-1999: 99US-0150884  
 PR 27-AUG-1999: 99US-0151065  
 PR 27-AUG-1999: 99US-0151066  
 PR 27-AUG-1999: 99US-0151080  
 PR 30-AUG-1999: 99US-0151303  
 PR 31-AUG-1999: 99US-0151348  
 PR 01-SEP-1999: 99US-0151930  
 PR 07-SEP-1999: 99US-0152363  
 PR 10-SEP-1999: 99US-0153070  
 PR 13-SEP-1999: 99US-0154758  
 PR 15-SEP-1999: 99US-0154818  
 PR 16-SEP-1999: 99US-0154839  
 PR 20-SEP-1999: 99US-0154779  
 PR 22-SEP-1999: 99US-0155136  
 PR 23-SEP-1999: 99US-0155486  
 PR 24-SEP-1999: 99US-0155659  
 PR 28-SEP-1999: 99US-0156458  
 PR 28-SEP-1999: 99US-0156466  
 PR 04-OCT-1999: 99US-0157117  
 PR 05-OCT-1999: 99US-0157753  
 PR 06-OCT-1999: 99US-0157865  
 PR 07-OCT-1999: 99US-0158029  
 PR 08-OCT-1999: 99US-0158242  
 PR 12-OCT-1999: 99US-0158369  
 PR 13-OCT-1999: 99US-0158493  
 PR 13-OCT-1999: 99US-0158494  
 PR 13-OCT-1999: 99US-0158495  
 PR 14-OCT-1999: 99US-0159329  
 PR 14-OCT-1999: 99US-0159430  
 PR 14-OCT-1999: 99US-0159431  
 PR 14-OCT-1999: 99US-0159437  
 PR 18-OCT-1999: 99US-0159684  
 PR 21-OCT-1999: 99US-0160741  
 PR 21-OCT-1999: 99US-0160767

PR 21-OCT-1999: 99US-0160768  
 PR 21-OCT-1999: 99US-0160770  
 PR 21-OCT-1999: 99US-0160814  
 PR 21-OCT-1999: 99US-0160815  
 PR 21-OCT-1999: 99US-0160880  
 PR 21-OCT-1999: 99US-0160981  
 PR 21-OCT-1999: 99US-0160989  
 PR 25-OCT-1999: 99US-0161404  
 PR 25-OCT-1999: 99US-0161405  
 PR 25-OCT-1999: 99US-0161406  
 PR 25-OCT-1999: 99US-0161359  
 PR 25-OCT-1999: 99US-0161360  
 PR 25-OCT-1999: 99US-0161361  
 PR 25-OCT-1999: 99US-0161362  
 PR 28-OCT-1999: 99US-0161992  
 PR 28-OCT-1999: 99US-0161993  
 PR 29-OCT-1999: 99US-0162142

Query Match 1.5% Score 8: DB 21: Length 318:  
 Best Local Similarity 100.0% Field No. 27:  
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Caps 0:

QY 491 DAKFKEL 498  
 DB 72 DAKFKEL 79

RESULT 15  
 ID AAC41937 standard: Protein: 372 AA.  
 AC AAC41937:  
 XX  
 XX 18-OCT-2000 (first entry)  
 XX  
 XX Arabidopsis thaliana protein fragment Sbj ID NO: 52239.  
 DE  
 XX Protein identification, signal transduction pathway, metabolic pathway;  
 KM hybridization assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 XX  
 XX Arabidopsis thaliana.  
 OS  
 XX  
 XX P1033405-A2.  
 FD  
 XX 06-SEP-2000.  
 XX  
 XX 25-FEB-2000: 2000EP-0201439  
 XX  
 PR 25-FEB-1999: 99US-0121825  
 PR 05-MAY-1999: 99US-0123180  
 PR 05-MAY-1999: 99US-0123448  
 PR 25-MAY-1999: 99US-0125788  
 PR 26-MAY-1999: 99US-0125789  
 PR 01-AUG-1999: 99US-0126785  
 PR 05-AUG-1999: 99US-0127472  
 PR 05-AUG-1999: 99US-0128234  
 PR 05-AUG-1999: 99US-0128714  
 PR 15-AUG-1999: 99US-0129645  
 PR 19-AUG-1999: 99US-0130077  
 PR 21-AUG-1999: 99US-0130449  
 PR 21-AUG-1999: 99US-0130510  
 PR 23-AUG-1999: 99US-0130891  
 PR 28-APR-1999: 99US-0131449  
 PR 30-APR-1999: 99US-0132448  
 PR 30-APR-1999: 99US-0132407  
 PR 04-MAY-1999: 99US-0132484  
 PR 05-MAY-1999: 99US-0132485  
 PR 05-MAY-1999: 99US-0132486  
 PR 07-MAY-1999: 99US-0132487  
 PR 07-MAY-1999: 99US-0132863  
 PR 11-MAY-1999: 99US-0134256  
 PR 14-MAY-1999: 99US-0134218

PR 14 MAY-1999; 990S-0134219  
PR 14 MAY-1999; 990S-0134221  
PR 14 MAY-1999; 990S-0134270  
PR 18 MAY-1999; 990S-0134768  
PR 19 MAY-1999; 990S-0134941  
PR 20 MAY-1999; 990S-0135124  
PR 21 MAY-1999; 990S-0135353  
PR 24 MAY-1999; 990S-0135659  
PR 25 MAY-1999; 990S-0136021  
PR 27 MAY-1999; 990S-0136392  
PR 28 MAY-1999; 990S-0136782  
PR 01-JUN-1999; 990S-0137222  
PR 03-JUN-1999; 990S-0137528  
PR 04-JUN-1999; 990S-0137502  
PR 07-JUN-1999; 990S-0137724  
PR 08-JUN-1999; 990S-0138094  
PR 10-JUN-1999; 990S-0138540  
PR 10-JUN-1999; 990S-0138847  
PR 14-JUN-1999; 990S-0139119  
PR 16-JUN-1999; 990S-0139452  
PR 16-JUN-1999; 990S-0139452  
PR 17-JUN-1999; 990S-0139493  
PR 18-JUN-1999; 990S-0139454  
PR 18-JUN-1999; 990S-0139455  
PR 18-JUN-1999; 990S-0139456  
PR 18-JUN-1999; 990S-0139457  
PR 18-JUN-1999; 990S-0139458  
PR 18-JUN-1999; 990S-0139459  
PR 18-JUN-1999; 990S-0139460  
PR 18-JUN-1999; 990S-0139461  
PR 18-JUN-1999; 990S-0139462  
PR 18-JUN-1999; 990S-0139463  
PR 18-JUN-1999; 990S-0139750  
PR 18-JUN-1999; 990S-0139763  
PR 21-JUN-1999; 990S-0139817  
PR 22-JUN-1999; 990S-0139899  
PR 23-JUN-1999; 990S-0140353  
PR 23-JUN-1999; 990S-0140354  
PR 24-JUN-1999; 990S-0140645  
PR 28-JUN-1999; 990S-0140823  
PR 29-JUN-1999; 990S-0140991  
PR 30-JUN-1999; 990S-0141287  
PR 01-JUL-1999; 990S-0141842  
PR 01-JUL-1999; 990S-0142154  
PR 02-JUL-1999; 990S-0142065  
PR 06-JUL-1999; 990S-0142390  
PR 08-JUL-1999; 990S-0142803  
PR 09-JUL-1999; 990S-0142920  
PR 12-JUL-1999; 990S-0142977  
PR 13-JUL-1999; 990S-0143542  
PR 14-JUL-1999; 990S-0143624  
PR 15-JUL-1999; 990S-0144005  
PR 16-JUL-1999; 990S-0144085  
PR 16-JUL-1999; 990S-0144086  
PR 19-JUL-1999; 990S-0144325  
PR 19-JUL-1999; 990S-0144331  
PR 19-JUL-1999; 990S-0144342  
PR 19-JUL-1999; 990S-0144343  
PR 19-JUL-1999; 990S-0144334  
PR 19-JUL-1999; 990S-0144335  
PR 20-JUL-1999; 990S-0144352  
PR 20-JUL-1999; 990S-0144632  
PR 20-JUL-1999; 990S-0144684  
PR 21-JUL-1999; 990S-0144814  
PR 21-JUL-1999; 990S-0145086  
PR 21-JUL-1999; 990S-0145088  
PR 22-JUL-1999; 990S-0145085  
PR 22-JUL-1999; 990S-0145087  
PR 22-JUL-1999; 990S-0145089  
PR 22-JUL-1999; 990S-0145192  
PR 23-JUL-1999; 990S-0145218  
PR 23-JUL-1999; 990S-0145218  
PR 23-JUL-1999; 990S-0145224

---

PR 26-JUL-1999; 990S-0145976  
PR 27-JUL-1999; 990S-0145973  
PR 27-JUL-1999; 990S-0145978  
PR 27-JUL-1999; 990S-0145979  
PR 28-JUL-1999; 990S-0145951  
PR 02-AUG-1999; 990S-0146386  
PR 02-AUG-1999; 990S-0146389  
PR 02-AUG-1999; 990S-0146388  
PR 03-AUG-1999; 990S-0147204  
PR 04-AUG-1999; 990S-0147204  
PR 04-AUG-1999; 990S-0147302  
PR 05-AUG-1999; 990S-0147192  
PR 05-AUG-1999; 990S-0147260  
PR 06-AUG-1999; 990S-0147303  
PR 06-AUG-1999; 990S-0147416  
PR 06-AUG-1999; 990S-0147443  
PR 09-AUG-1999; 990S-0147945  
PR 10-AUG-1999; 990S-0148171  
PR 11-AUG-1999; 990S-0148319  
PR 12-AUG-1999; 990S-0148341  
PR 13-AUG-1999; 990S-0148665  
PR 13-AUG-1999; 990S-0148684  
PR 16-AUG-1999; 990S-0149368  
PR 17-AUG-1999; 990S-0149175  
PR 18-AUG-1999; 990S-0149426  
PR 20-AUG-1999; 990S-0149722  
PR 20-AUG-1999; 990S-0149723  
PR 20-AUG-1999; 990S-0149929  
PR 23-AUG-1999; 990S-0149902  
PR 23-AUG-1999; 990S-0149930  
PR 25-AUG-1999; 990S-0150566  
PR 25-AUG-1999; 990S-0150844  
PR 27-AUG-1999; 990S-0151065  
PR 27-AUG-1999; 990S-0151066  
PR 27-AUG-1999; 990S-0151080  
PR 30-AUG-1999; 990S-0151303  
PR 31-AUG-1999; 990S-0151438  
PR 01-SEP-1999; 990S-0151930  
PR 07-SEP-1999; 990S-0152363  
PR 10-SEP-1999; 990S-0153070  
PR 13-SEP-1999; 990S-0153758  
PR 15-SEP-1999; 990S-0154018  
PR 16-SEP-1999; 990S-0154018  
PR 20-SEP-1999; 990S-0154779  
PR 20-SEP-1999; 990S-0155139  
PR 23-SEP-1999; 990S-0155486  
PR 24-SEP-1999; 990S-0155659  
PR 28-SEP-1999; 990S-0156458  
PR 29-SEP-1999; 990S-0156566  
PR 04-OCT-1999; 990S-0157117  
PR 05-OCT-1999; 990S-0157753  
PR 06-OCT-1999; 990S-0157865  
PR 07-OCT-1999; 990S-0158029  
PR 08-OCT-1999; 990S-0158232  
PR 12-OCT-1999; 990S-0158266  
PR 13-OCT-1999; 990S-0159293  
PR 13-OCT-1999; 990S-0159294  
PR 13-OCT-1999; 990S-0159295  
PR 14-OCT-1999; 990S-0159329  
PR 14-OCT-1999; 990S-0159330  
PR 14-OCT-1999; 990S-0159331  
PR 14-OCT-1999; 990S-0159637  
PR 14-OCT-1999; 990S-0159638  
PR 18-OCT-1999; 990S-0159584  
PR 21-OCT-1999; 990S-0160741  
PR 21-OCT-1999; 990S-0160767  
PR 21-OCT-1999; 990S-0160768  
PR 21-OCT-1999; 990S-0160770  
PR 21-OCT-1999; 990S-0160814  
PR 21-OCT-1999; 990S-0160815  
PR 22-OCT-1999; 990S-0160980  
PR 22-OCT-1999; 990S-0160981  
PR 22-OCT-1999; 990S-0160989

PR 25-OCT-1999: 99US-0161404.  
PR 25-OCT-1999: 99US-0161405.  
PR 25-OCT-1999: 99US-0161406.  
PR 26-OCT-1999: 99US-0161359.  
PR 26-OCT-1999: 99US-0161360.  
PR 26-OCT-1999: 99US-0161361.  
PR 28-OCT-1999: 99US-0161920.  
PR 28-OCT-1999: 99US-0161992.  
PR 28-OCT-1999: 99US-0161993.  
PR 29-OCT-1999: 99US-0162142.

Query Match 1.5%; Score 8; DB 21; Length 372;  
Best Local Similarity 100.0%; Pred. No. 31;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 491 DAKTFKEL 498  
|||||||  
DB 126 DAKTFKEL 133

Search completed: July 15, 2003, 09:40:39  
Job time : 74 secs





A:Residues: 1-580 <TEP>  
 A:Cross-references: CR:AF002528; CR:AF002094; NID:q722734; PIRN:AA042120.1; PIR:q722703  
 A:Experimental source: serogroup A, strain MCS8  
 C:Genetics:  
 A:Gene: NMB1780

Query Match 1.58; Score 8; DB 2; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VNYDYNGK 414  
 |||||  
 DB 325 VNYDYNGK 332

RESULT 3  
 AB1989  
 probable periplasmic protein NMA0687 [imported] - Neisseria meningitidis (strain 22491)  
 C:Species: Neisseria meningitidis  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: AB1989  
 P:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,  
 J.; Holtroyd, S.; Jorgels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.  
 A:Reference number: AB1775; MIM:2032556; PMID:10761919  
 A:Accession: AB1989  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-580 <PAR>  
 A:Cross-references: CR:AL12773; CR:AL157959; NID:q7379120; FICN:CA681973.1; PIR:q737941  
 A:Experimental source: serogroup A, strain 22491  
 C:Genetics:  
 A:Gene: NMA0687; NMA0687

Query Match 1.58; Score 8; DB 2; Length 580;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VNYDYNGK 414  
 |||||  
 DB 325 VNYDYNGK 332

RESULT 4  
 158411  
 protein-tyrosine kinase (EC 2.7.1.112) btk - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 04-Feb-2000  
 C:Accession: 158411  
 R:Fujimoto, J.; Yamamoto, T.  
 Oncogene 9, 693-698, 1994  
 A:Title: btk, a mouse gene encoding a novel receptor type protein-tyrosine kinase, is put  
 A:Reference number: 158411; MIM:194150990; PMID:8108111  
 A:Accession: 158411  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-856 <RPS>  
 A:Cross-references: CR:017383; NID:q801867; PIRN:RAA04216.1; PIR:q801868  
 C:Superfamily: protein-tyrosine kinase axl; fibronectin type III repeat homology; immun  
 C:Keywords: ATP, glycoprotein, phosphotransferase, tyrosine-specific protein kinase  
 F:142-191/Domain: immunoglobulin homology <IMM>  
 F:501-778/Domain: protein kinase homology <KIN>  
 F:509-517/Region: protein kinase ATP-binding motif

Query Match 1.58; Score 8; DB 2; Length 856;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGIALLHS 59  
 |||||  
 DB 236 DGIALLHS 243

RESULT 5  
 149152  
 protein-tyrosine kinase (EC 2.7.1.112) tyro3, isoform B - mouse

N:Alternate names: tyrosine kinase growth factor receptor  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 04-Feb-2000

C:Accession: 149152; 148861  
 R:Schultz, N.T.; Paulhac, C.L.; Lee, L.; Zhou, K.  
 Brain Res. Mol. Brain Res. 28, 278-280, 1995

A:Title: Isolation and expression analysis of tyro3, a murine growth factor receptor  
 A:Reference number: 149151; MIM:95240399; PMID:7723626  
 A:Accession: 149152

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-876 <RPS>

A:Cross-references: EMBL:U18343; NID:q687627; PIRN:AA02943.1; PIR:q687628  
 R:Riesacker, L.G.; Giamola, D.M.; Emerson, S.G.  
 Oncogene 10, 2239-2242, 1995

A:Title: Identification of alternative exons, including a novel exon, in the tyrosine  
 A:Reference number: 148861; MIM:9530487; PMID:7784089  
 A:Accession: 148861

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-88 <RPS>

A:Cross-references: EMBL:U23722; NID:q301332; PIRN:AA072415.1; PIR:q401994  
 C:Genetics:  
 A:Gene: tyro3; Etk2/tyro3

C:Superfamily: protein-tyrosine kinase axl; fibronectin type III repeat homology; imm  
 C:Keywords: ATP, glycoprotein, growth factor receptor, phosphotransferase, tyrosine-  
 F:43-105/Domain: immunoglobulin homology <IMM>  
 F:308-391/Domain: fibronectin type III repeat homology <3FR>  
 F:502-779/Domain: protein kinase homology <KIN>  
 F:510-518/Region: protein kinase ATP-binding motif

Query Match 1.58; Score 8; DB 2; Length 876;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGIALLHS 59  
 |||||  
 DB 236 DGIALLHS 243

RESULT 6  
 J04166  
 protein-tyrosine kinase (EC 2.7.1.112) tyro3 precursor - rat  
 N:Alternate names: protein-tyrosine kinase src; receptor-type tyrosine kinase lse  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: J04166; P0185  
 R:Ohashi, K.; Henda, S.; Ichimura, Y.; N. Nakamura, I.; Mizuno, K.  
 J. Biochem. 117, 1267-1275, 1995  
 A:Title: Molecular cloning and in situ localization in the brain of rat Src receptor  
 A:Reference number: J04166; MIM:64104999; PMID:7490270  
 A:Accession: J04166  
 A:Molecule type: mRNA  
 A:Residues: 1-880 <OHM>  
 A:Cross-references: PIRN:R47880; NID:q1494195; PIRN:RAA07114.1; PIR:q829057  
 A:Experimental source: brain  
 A:Note: It is uncertain whether Met-1 or Met-7 is the initiator  
 Relat. C.: Lemke, G.  
 Neuron 6, 691-704, 1991

A:Title: An extended family of protein tyrosine kinase genes differentially expressed  
 A:Reference number: P0183; MIM:6122566; PMID:2057425  
 A:Accession: P0185

A:Molecule type: mRNA  
 A:Residues: 650-703 <LAI>  
 A:Experimental source: sciatic nerve  
 C:Comment: This receptor plays an important role in development, function, and mainte  
 C:Genetics:  
 A:Gene: tyro-3

C:Superfamily: protein-tyrosine kinase axl; fibronectin type III repeat homology; immunoglobulin domain; growth factor receptor; phosphotransferase; transforming  
 C:Keywords: ATP, glycoprotein, growth factor receptor, phosphotransferase, transforming  
 F:1-31/Domain: (or 7-31) signal sequence #status predicted <SIG>  
 F:32-880/Product: (or 32-874) protein-tyrosine kinase tyros3 #status predicted <MAT>  
 F:50-109/Domain: immunoglobulin homology <IMM1>  
 F:146-195/Domain: immunoglobulin homology <IMM2>  
 F:214-293/Domain: fibronectin type III repeat homology <FN3A>  
 F:312-395/Domain: fibronectin type III repeat homology <FN3B>  
 F:419-441/Domain: transmembrane #status predicted <TM>  
 F:506-783/Domain: protein kinase homology <KIN>  
 F:514-522/Region: protein kinase ATP-binding motif  
 F:515-766/Domain: cytoplasmic tyrosine kinase #status predicted <CTK>  
 F:53,75,181,220,230,283,356,370/Binding site: catalytic site (Asn) (covalent) #status pred

Query Match 1.5% Score 8, DB 1, Length 880,  
 Best Local Similarity 100.0% Pred No 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
 |||||  
 Db 240 DGLALLHS 247

RESULT 7  
 B53743  
 protein-tyrosine kinase (EC 2.7.1.112) tyros3 precursor mouse  
 N:Alternate names: protein-tyrosine kinase sky; receptor-type tyrosine kinase rse; tyros  
 C:Species: Mus musculus (house mouse)  
 C:Date: 07-Oct-1994 #sequence-revision 07-Oct-1994 #text-change 18-Jun-1999  
 C:Accession: B53743; S44141; J02146; I49198; I49151; I48863; I48860  
 J:Mark: M.P.: Scadden, D.T.: Wang, Z.: Gu, Q.: Goddard, A.: Godowski, P.J.  
 R: Biol. Chem. 269, 10720-10728, 1994  
 A>Title: rse, a novel receptor-type tyrosine kinase with homology to Ax1/Oto, is express  
 A:Reference number: A57743; MIM:94193774; PMID:7511603  
 A:Accession: B53743  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-880 <MAR>  
 A:Cross-references: GI:105584; NID:9485518; P10N:AAA19237 1; P10:9485519  
 R:Alali, C.; Gore, M.; Lemke, G.  
 Submitted to the EMBL Data Library, March 1994  
 A:Description: Structure, expression and activity of TYRO3.  
 A:Reference number: S44141  
 A:Accession: S44141  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 7-629, 647, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1685, 1686, 1687, 1688, 1689, 1690, 1691, 1692, 1693, 1694, 1695, 1696, 1697, 1698, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1715, 1716, 1717, 1718, 1719, 1720, 1721, 1722, 1723, 1724, 1725, 1726, 1727, 1728, 1729, 1730, 1731, 1732, 1733, 1734, 1735, 1736, 1737, 1738, 1739, 1740, 1741, 1742, 1743, 1744, 1745, 1746, 1747, 1748, 1749, 1750, 1751, 1752, 1753, 1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768, 1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783, 1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798, 1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559

hypothetical protein APE2498 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Apr-1999 #text\_change 20-Jun-2000  
 C:Accession: H72482  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamizawa, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, S.; K DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic ctenarchaeon, Aeropyrum pernix strain K1  
 A:Reference number: A72450; MUID:99310339; PMID:10382966  
 A:Accession: H72482  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-111 <KAW>  
 A:Cross-references: DDBJ:AB000064; NID:95105945; P10N:HA01514.1; P1D:95106203  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE2498  
 C:Superfamily: Aeropyrum pernix hypothetical protein APE2498

Query Match 1.38; Score 7; DB 2; Length 111,  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 165 AVNLATS 171  
 |||||  
 DB 35 AVNLATS 41

RESULT 10  
 T05467  
 hypothetical protein T805.60 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Apr-2000  
 C:Accession: T05467  
 R:Hevan, M.; Medler, H.; Wambutt, R.; Hancock, J.; Mewes, H.W.; Mayor, K.F.X.; Schmolze submitted to the Protein Sequence Database, February 1998  
 A:Reference number: Z15417  
 A:Accession: T05467  
 A:Molecule type: DNA  
 A:Residues: 1-124 <HEV>  
 A:Cross-references: EMBL:AL021890  
 A:Experimental source: cultivar Columbia; EAC clone T805  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 40/2; 92/3  
 A:Note: T805.60  
 C:Superfamily: hypothetical protein YC1033c

Query Match 1.38; Score 7; DB 2; Length 124,  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 378 PITSATV 384  
 |||||  
 DB 2 PITSATV 8

RESULT 11  
 B87322  
 chemotaxis protein CheY [Imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: B87322  
 R:Kernan, M.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Debby, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolton, D.J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M DNA Res. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; MUID:21173698; PMID:11256647  
 A:Accession: B87322  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-125 <STO>

A:Cross-references: GB:AE035673; NID:913421788; P10N:AK22574.1; GSP08:GN0148  
 C:Genetics:  
 A:Gene: C030588  
 C:Superfamily: chemotaxis cheY protein, response regulator homolog

Query Match 1.38; Score 7; DB 2; Length 125,  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 52 DGLALLH 58  
 |||||  
 DB 61 DGLALLH 67

RESULT 12  
 H71197  
 hypothetical protein PH1854 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 28-Jul-2000  
 C:Accession: H71197  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se M.; Ohjima, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, S.; Kishida, N.; Ogi DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A:Reference number: A71000; MUID:98344137; PMID:9679194  
 A:Accession: H71197  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-140 <KAW>  
 A:Cross-references: GB:AP000007; NID:9736134; P10N:HA10975.1; P1D:97362792  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an earlier accession for a sequence replaced by Genba C:Genetics:  
 A:Gene: PH1854  
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1854

Query Match 1.38; Score 7; DB 2; Length 140,  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 347 NKLTSP 353  
 |||||  
 DB 13 NKLTSP 19

RESULT 13  
 T35580  
 probable acetyltransferase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jul-2000  
 C:Accession: T35580  
 R:Saunders, D.; Harris, D.; Parkhill, J.; Barrett, B.G.; Kalandroum, M.A. submitted to the EMBL Data Library, August 1998  
 A:Reference number: Z21583  
 A:Accession: T35580  
 A:Status: preliminary; translated from GR/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-177 <SAU>  
 A:Cross-references: EMBL:AF031317; P10N:CAA20407.1; GSP08:GN0070; SGP08:SG0624.29  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SGP08:SG0624.29  
 C:Superfamily: Escherichia coli ribosomal protein-alanine N-acetyltransferase that

Query Match 1.38; Score 7; DB 2; Length 177,  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 313 EEDGHV 319  
 |||||  
 DB 63 EEDGHV 69



## RESULT 14

G69222

molybdenum formylmethanofuran dehydrogenase, subunit E Methanobacterium thermoautotroph

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Oct-1999

C:Accession: G69222

R:Smith, D.R.; Donnerette-Stamm, L.A.; DeLong-Eberly, C.; Lee, H.; Dubois, J.; Aldridge, T.;

Qiu, D.; Spadafino, R.; Voth, R.; Wang, Y.; Wierzbowski, J.; Gilson, P.; Jiwani, N.

K.; Church, G.M.; Daniels, C.J.; Mo, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func

A:Reference number: A63000, MIMD:98037514, PMID:9371463

A:Accession: G69222

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-180 &lt;MTH&gt;

A:Cross-references: GB:AE000867; GB:AE000868; NID:9222009; P10N:AA09445.1; P1D:926201

A:Experimental source: strain Delta H

A:Gene: MTH917

A:Start codon: GTG

C:Superfamily: Methanosarcina barkeri molybdenum formylmethanofuran dehydrogenase chain

Query Match 1.3% Score 7; DB 2; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 VPEKKEK 211

|||||

Db 147 VPEKKEK 153

## RESULT 15

T49476

related to 50S RIBOSOMAL PROTEIN L2 [imported] - Neurospora crassa

N:Alternate names: protein B14D6.280

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49476

R:Schulze, H.; Altmann, V.; Hehlsel, J.; Brandt, P.; Farlmann, R.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: 225022

A:Accession: T49476

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1181 &lt;SCH&gt;

A:Cross-references: EMBL AL356173; GSPPE 3M60116; NCSP: B14D6.280

A:Experimental source: BAC clone B14D6; strain OR74A

C:Genetics:

A:Gene: NCSP:B14D6.280

A:Map position: 6

Query Match 1.3% Score 7; DB 2; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 EKEKKKS 213

|||||

Db 108 EKEKKKS 114

Search completed: July 15, 2003, 09:49:19  
 Job time : 45 secs



Downloaded from <http://ajph.org/> on November 10, 2015

2000

Db 151 NDLGAFTE 158

```

RESULT 2
TYO3_MOUSE
ID TYO3_MOUSE STANDARD: PRT: 880 AA
AC P55146
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase receptor TYRO3 precursor (P1.2.7.1.112)
DE (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase DTK) (TK19-2).
CN TYRO3 OR ETK OR RSE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA MEDLINE=9419374; PubMed=7511603;
RX Mark M.R., Scadden D.T., Wang Z., Gu Q., Goddard A., Godowski P.J.;
RT "RSE, a novel receptor-type tyrosine kinase with homology to Axl/Uto,
RT is expressed at high levels in the brain.";
RL J. Biol. Chem. 269:10720-10728(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA MEDLINE=95161079; PubMed=7857657;
RX Crosier P.S., Lewis P.M., Hall L.R., Vitas M.R., Morris C.M.,
RA Beier D.R., Wood C.R., Crosier K.R.;
RT "Isolation of a receptor tyrosine kinase (DTK) from embryonic stem
RT cells: structure, genetic mapping and analysis of expression.";
RL Growth Factors 11:125-136(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=94336210; PubMed=8058320;
RA Lai C., Gore M., Lemke G.;
RT "Structure, expression, and activity of Tyro 3, a neural adhesion-
RT related receptor tyrosine kinase.";
RL Oncogene 9:2567-2578(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c; TISSUE-Brain;
RX MEDLINE=95240399; PubMed=7723626;
RA Schulz N., Pauliac C., Lee L., Zhou R.;
RT "Isolation and expression analysis of Tyro3, a murine growth factor
RT receptor tyrosine kinase preferentially expressed in adult brain";
RL Brain Res Mol. Brain Res. 28:273-280(1995).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c; TISSUE-Brain;
RA Sasaki M.;
RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RT PUNCTION: MAY BE INVOLVED IN CELL ADHESION PROCESSES, PARTICULARLY
RT IN THE CENTRAL NERVOUS SYSTEM.
CC -1 CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC -1 TISSUE SPECIFICITY: ABUNDANT IN THE BRAIN AND LOWER LEVELS IN
CC OTHER TISSUES.
CC -1 SIMILARITY: RELATES TO THE TYR FAMILY OF PROTEIN KINASES.
CC AXI/UTO SUBFAMILY.
CC -1 SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1 SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U05683; AAA19237.1;
DR EMBL: U18933; AAC52148.1;
DR EMBL: X78103; CA834995.1; ALT_INT.
DR EMBL: U18342; AAB26942.1; ALT_INT.
DR EMBL: AB000828; BAA19193.1;
DR HSSP: P11362; IFGK.
DR MCD: M01104294; TYRO3.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR001245; Ig_T_kinase.
DR Pfam: PF00047; Irc3_2.
DR Pfam: PF00064; Irc3_2.
DR Pfam: PF00064; Irc3_2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00410; Ig_Like; 1.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00219; TYRO3; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Receptor: Glycoprotein, Tyrosine-protein kinase, ATP-binding;
KW Transferase; Phosphorylation; Transmembrane; Signal; Repeat;
KW Immunoglobulin domain
FT SIGNAL 1 30
FT CHAIN 1 880
FT DOMAIN 31 419
FT TRANSMM 420 440
FT DOMAIN 441 880
FT DOMAIN 47 114
FT DOMAIN 143 209
FT DOMAIN 214 301
FT DOMAIN 305 401
FT DOMAIN 508 785
FT NP_BIND 514 522
FT BINDING 540 540
FT ACT_SITE 645 645
FT DISULF 54 107
FT DISULF 150 193
FT CARBOHYD 53 53
FT CARBOHYD 75 75
FT CARBOHYD 181 181
FT CARBOHYD 220 220
FT CARBOHYD 230 230
FT CARBOHYD 283 283
FT CARBOHYD 356 356
FT CARBOHYD 370 370
FT MOF_PES 676 676
FT CONFLICT 630 630
FT CONFLICT 811 811
SQ SEQUENCE 880 AA; 9622 MW; 3B7AC563E01B347 CPE64;
Query Match 1.58; Score 8; DB 1; Length 880;
Best Local Similarity 100.0%; Prod. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 52 DGTALLHS 59
DB 240 DGTALLHS 247
RESULT 3
TYO3_MOUSE
ID TYO3_MOUSE STANDARD: PRT: 880 AA
AC P55146
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase receptor TYRO3 precursor (P1.2.7.1.112)
DE (Tyrosine-protein kinase SKY).

```



UT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Hypothetical tRNA/tRNA methyltransferase yjH (EC 2.1.1.-)  
 GN YPH OR B4180 OR Z5787 OR EC55156.  
 OS Escherichia coli, and  
 CC Escherichia coli O157:H7.  
 CC Bacteria: Proteobacteria, gamma subdivision, Enterobacteriaceae:  
 CC Escherichia.  
 CC NCBI\_TaxID=562, 83334.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=95334362, PubMed-7610340;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes."  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed-11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.T., Davis N.W., Lim A., Diallanta E.T., Potamocis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
 RL Nature 409:529-533(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kobayashi S., Shiba T., Hattori M., Shinagawa H.,  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12."  
 RL DNA Res. 8:11-22(2001).  
 CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRPM FAMILY.  
 CC -1- STRONG: TO H INFLUENZAE H10860.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: U14003; AAA97076.1;  
 DR EMBL: AE000490; AAC77137.1;  
 DR EMBL: AE000551; AAG59476.1;  
 DR EMBL: AF002568; BAB38579.1;  
 DR EMBL: E312483; YjH.  
 DR ECGene: E312483; YjH.  
 DR InterPro: IPR001537; SpoM\_methylase.  
 DR InterPro: IPR004441; RNA\_methyl\_13.  
 DR Pfam: PF00568; SpoM\_methylase\_1.  
 DR ProDom: PD001243; SpoM\_methylase\_1.  
 DR TrpPfam: T19P00186; rRNA\_methyl\_3\_1.  
 DR TIGPfam: T19P00186; rRNA\_methyl\_3\_1.  
 KM Hypothetical protein; Transferase; Methyltransferase;  
 KW Complete proteome.  
 SU SPOUNCE 243 AA; 26556 MW; 606417054034106 Cp664;  
 Query Match 1.38; Score 7; DB 1; Length 243.  
 Host Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 ID YSCY\_YERPE STANDARD; PRI; 261 AA.  
 AC P40299;  
 FT 01-FEB-1995 (Rel. 41, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Yop proteins translocation protein T.  
 GN YSCY OR YPCD1.46 OR Y0035.  
 OS Yersinia pestis, and  
 OS Yersinia pseudotuberculosis.  
 CC Plasmid pCD1, and Plasmid pIR1.  
 CC Bacteria, Proteobacteria, gamma subdivision; Enterobacteriaceae:  
 CC Yersinia.  
 CC NCBI\_TaxID=632, 633;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Y.pestis; STRAIN-KIM5; PLASMID-pCD1;  
 RX MEDLINE=98427122; PubMed-9746557;  
 RA Perty P.D., Straloy S.C., Robertson J.D., Rose D.J., Gregor J.,  
 RA Blattner F.R.;  
 RT "RNA sequencing and analysis of the low G+C response plasmid pCD1 of  
 RT Yersinia pestis KIM5."  
 RL Infect Immun 66:4611-4623(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Y.pestis; STRAIN-KIM5; PLASMID-pCD1;  
 RX MEDLINE=98422474; PubMed-9748454;  
 RA Hu P., Elliott J., McCreedy P., Skowronski E., Barnes J.,  
 RA Kobayashi A., Prubaker R.R., Garcia E.;  
 RT "Structural organization of virulence-associated plasmids of Yersinia  
 RT pestis."  
 RL J Bacteriol 180:5192-5202(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Y.pestis; STRAIN-KIM5; PLASMID-pIR1;  
 RX MEDLINE=21470413; PubMed-11586460;  
 RA Parkhill J., Wren B.W., Thomson N.R., Tittall R.W., Holden M.T.G.,  
 RA Penrice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Gotoh-Oharana A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Fellwell T., Hamlin N., Holtroyd S., Jorgensen K., Kallsgaard A.V.,  
 RA Leather S., Moutle S., Oyston P.C.P., Quail M., Rutherford K.,  
 RA Saunders M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RT "Genome sequence of Yersinia pestis: the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Y.pestis; STRAIN-YPIII; PLASMID-pIR1;  
 RX MEDLINE=94222840; PubMed-8109210;  
 RA Bergman T., Erickson K., Galyov E., Persson C., Wolf-Watz H.;  
 RT "The yscY (yscN) gene cluster of Yersinia pseudotuberculosis is  
 RT involved in Yop secretion and shows high homology to the spa gene  
 RT clusters of Shigella flexneri and Salmonella typhimurium."  
 RL J Bacteriol 176:2619-2626(1994).  
 CC -1- FUNCTION: COMPONENT OF THE YOP SECRETION MACHINERY.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -1- SIMILARITY: BELONGS TO THE FILH/McW/SPAR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AF074612; AAC69785.1;  
 DR EMBL: AF053946; AAC62558.1;  
 DR EMBL: AL117189; CAH54923.1;  
 DR EMBL: L25667; AAA27680.1;  
 DR InterPro: IPR002010; Bac\_export\_1.

```

DR Pfam: PF01311: Bac.exp0rt_1; 1.
DR PRINTS: PR00953; TYPE3IMPROT.
KW Transport; Protein transport; Plasmid; Virulence; Transmembrane;
Complete proteome.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
SQ SEQUENCE 261 AA; 28451 MW; 1EE3BE9E07AD1F3 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GVLRNG 36
DB 37 GVLRNG 43

RESULT 7
ADD_RHIL0 STANDARD: PRT; 324 AA
AC 098GV2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Adenosine deaminase (EC 3.5.4.4) (Adenosine aminohydrolase).
GN ADD OR ML13163.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA MEDLINE=21082930; PubMed=11214968.
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,
RA Takuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti".
RL DNA Res. 7:331-338(2000).
CC -1- CATALYTIC ACTIVITY: Adenosine + H(2)O = inosine + NH(3).
CC -1- SIMILARITY: RELINGS TO THE ADENOSINE AND AMP DEAMINASE FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP003001; BAB50114.1; -
DR InterPro: IPR001365; A/AMP_deaminase.
DR Pfam: PF00962; A_deaminase; 1.
KW Hydrolyase; Nucleotide metabolism; Complete proteome.
FT ACT_SITE 186 186 POTENTIAL.
FT ACT_SITE 234 234 POTENTIAL.
FT ACT_SITE 267 267 POTENTIAL.
FT ACT_SITE 268 268 POTENTIAL.
SQ SEQUENCE 324 AA; 35247 MW; 241959E4F90EA CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 FVDRKTK 291
DB 306 FVDRKTK 312

RESULT 8
TRMA_NEIMB STANDARD: PRT; 362 AA.
ID TRMA_NEIMB
AC 09JYA0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE tRNA (Uracil-5'-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54)-
DE methyltransferase) (K0MT).
GN TRMA OR NMB1679.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
PC STPAIN-M758 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickory F.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty R.A.,
RA Mason T., Gleckler A., Parksey P.S., Blair E., Cline H., Clark E.B.,
RA Cotton M.D., Hitterberg T.R., Khouf H., Qin H., Vamathevan J.,
RA Gyll T., Scharf V., Mastigiani V., Pizarro M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.P., Rappaport R., Venter I.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58".
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
CC 54 (M-5-U54) in all tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing thymine.
CC -1- SIMILARITY: RELINGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRMA
CC SUBFAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF002518; AAF42927.1; -
DR TIGR: NMB1679; -
DR InterPro: IPR000051; SAM_bind.
DR ProSITE: PS01240; TRMA_1; 1.
DR ProSITE: PS01231; TRMA_2; FALSE_NEG.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 208 214 S-ADENOSYLMETHIONINE BINDING (BY
FT SIMILARITY).
FT ACT_SITE 318 318 BY SIMILARITY.
SQ SEQUENCE 362 AA; 41309 MW; 4CF18AF93E72F8EB CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 362;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 NIEANRI 88
DB 243 NIEANRI 249

RESULT 9
YGB1_METJA STANDARD: PRT; 380 AA.
ID YGB1_METJA
AC 059075;

```

DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein M1681.  
 OS M1681.  
 ON Methanococcus jannaschii.  
 OC Archaea, Euryarchaeota: Methanococci: Methanococcales;  
 OC Methanocaldococcaceae: Methanocaldococcus.  
 OX NCBI\_TaxID=2190.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Buit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton P.A., Gocayne J.P.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Peich C.I.,  
 RA Overbeek R., Kirkness F.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Uitterback L.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT \*Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.\*  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: TO M.THERMOPHILICUM MTH1684.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: 067608; AAB99702.1;  
 DR TIGR: M1681;  
 DR InterPro: IPR001450; 4Peas\_Ferredoxin.  
 DR InterPro: IPR002708; DUF39.  
 DR Pfam: PF00037; Icr4; 2.  
 DR Pfam: PF01837; DUF39; 1.  
 DR ProDom: PD011569; DUF39; 1.  
 DR Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 380 AA; 42580 MW; 03b2c17979f4746a CRC64;  
 Query Match 138; Score 7; Tr 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3/;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 336 LKRIKDD 342  
 DB 253 LKRIKDD 259  
 RESULT 10  
 CXX1\_MOUSE STANDARD; PFT: 381 AA.  
 AC P23242;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Gap junction alpha-1 protein (Connexin 43) (Cx43) (Gap junction 43 kDa  
 DE heart protein).  
 DR GJA1 OR CXN-43.  
 GN GJA1 OR CXN-43.  
 OS Mus musculus (Mouse).  
 OC Eukaryota, Metazoa, Chordata, Vertebrata: Euteleostomi;  
 OC Mammalia, Eutheria, Eutheria: Sciuromorphi: Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=91217014; PubMed=1708769;  
 RA Beyer E.C., Steinberg T.H.;  
 RT \*Evidence that the gap junction protein connexin-43 is the  
 RT ATP-induced pore of mouse macrophages.\*

RL J. Biol. Chem. 266:7971-7974(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Padayaratnam C.P., Morgan J.L., Lo C.W.;  
 RT Submitted (SEP 1991) to the EMBL/Genbank/Trna databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RX MEDLINE=9126228; PubMed=906097;  
 RT Nishi M., Kumar N.M., Gilula N.B.;  
 RT \*Developmental regulation of gap junction gene expression during  
 RT mouse embryonic development.\*  
 RL Dev. Biol. 146:117-130(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=9229686; PubMed=1318884;  
 RA Hennemann J., Stoklyna P., Liechtenborg-Frater H., Jungbluth S.,  
 RA Dahl E., Schwarz J., Nicholson B.J., Willock K.;  
 RT \*Molecular cloning and functional expression of mouse connexin40, a  
 RT second gap junction gene preferentially expressed in lung.\*  
 RL J. Cell Biol. 117:1299-1310(1992).  
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED  
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXINS, THROUGH WHICH  
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.  
 CC -1- FUNCTION: CONNEXIN 43 IS POSSIBLY THE ATP-INDUCED PORE OF  
 CC MOUSE MACROPHAGES.  
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. ALPHA-TYPE (GROUP 1)  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M61896; AAA37444.1;  
 DR EMBL: X61576; CAA43778.1;  
 DR EMBL: M61861; AAA3027.1;  
 DR EMBL: X62836; CAA4640.1;  
 DR PIR: A39802; A39802.  
 DR PIR: S18110; S18110.  
 DR PIR: S24110; S24110.  
 DR MGD: MG1:95713; G3a1.  
 DR InterPro: IPR000560; Connexin.  
 DR InterPro: IPR002261; Connexin43.  
 DR Pfam: PF00029; Connexin; 1.  
 DR Pfam: PF03508; Connexin43; 1.  
 DR PRINTS: PR00206; CONNEXIN.  
 DR SMART: SM00037; CNX; 1.  
 DR PROSITE: PS00407; CONNEXINS\_1; 1.  
 DR PROSITE: PS00408; CONNEXINS\_2; 1.  
 DE Gap junction; Transmembrane.  
 FT INTRAMET 0  
 FT DOMAIN 1 22  
 FT TRANSMEM 23 43  
 FT DOMAIN 44 75  
 FT TRANSMEM 76 96  
 FT DOMAIN 97 154  
 FT TRANSMEM 155 175  
 FT DOMAIN 176 206  
 FT TRANSMEM 207 227  
 FT DOMAIN 228 381  
 FT CONFLICT 319 319 M -> F (IN REF. 3).  
 SV SEQUENCE 381 AA; 42673 MW; 96958a57979c1a7b5 CRC64;  
 Query Match 138; Score 7; Tr 1; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 503 VNMVEMHL 509  
 DB 120 VNMVEMHL 126

RESULT 11

CC XA1\_RAT STANDARD: PRT: 381 AA.  
 AC P08050;  
 DT 01-AUG-1988 (rel. 08, created)  
 DT 01-NOV-1988 (rel. 09, last sequence update)  
 DT 16-OCT-2001 (rel. 40, last annotation update)  
 DE Gap junction alpha-1 protein (connexin 43) (Cx43) (gap junction 43 kDa heart protein).  
 GN GJA1 OR CXN-43.  
 OS Rattus norvegicus (Rat)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88047496; PubMed=2826492;  
 RA Beyer E.C., Paul D.L., Goodenough P.A.;  
 RT "Connexin43: a protein from rat heart homologous to a gap junction protein from liver."  
 RL J. Cell Biol. 105:2621-2629(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=uterus;  
 RC MEDLINE=91241281; PubMed=1852114;  
 RA Lang L.M., Beyer E.C., Schwartz A.L., Gillin J.D.;  
 RT "Molecular cloning of a rat uterine gap junction protein and analysis of gene expression during gestation."  
 RL Am. J. Physiol. 260:E787-E793(1991).  
 RN [3]  
 RP SEQUENCE OF 1-32.  
 RC TISSUE=Heart;  
 RX MEDLINE=85307050; PubMed=2987225;  
 RA Nicholson R.J., Gros D.R., Kent S.R.H., Hood L.F., Revel J.-P.;  
 RT "The Mr 28,000 gap junction proteins from rat heart and liver are different but related."  
 RL J. Biol. Chem. 260:6514-6517(1985).  
 RN [4]  
 RP PARTIAL SEQUENCE OF 1-15.  
 RC TISSUE=brain;  
 RX MEDLINE=91348048; PubMed=1652440;  
 RA Dupont E., el Aoumari A., Fromaget C., Briand J.-C., Gros D.;  
 RT "Affinity purification of a rat-brain junctional protein, connexin 43."  
 RL Eur. J. Biochem. 200:263-270(1991).  
 RN [5]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=91337077; PubMed=1651718;  
 RA John S.A., Revel J.-P.;  
 RT "Connexon integrity is maintained by non-covalent bonds: intramolecular disulfide bonds link the extracellular domains in rat connexin-43."  
 RL Biochem. Biophys. Res Commun 178:1312-1318(1991).  
 RN [6]  
 RP TOPOLOGY  
 RX MEDLINE=92167270; PubMed=1371548;  
 RA Yeager M., Gilula N.B.;  
 RT "Membrane topology and quaternary structure of cardiac gap junction ion channels."  
 RL J. Mol. Biol. 223:929-948(1992)  
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED PAIRS OF TRANSMEMBRANE CHANNELS. THE CONNEXONS, THROUGH WHICH MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.  
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- PTM: THERE IS AT LEAST ONE INTRAMOLECULAR DISULFIDE BOND  
 CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. ALPHA-TYPE (GROUP II)

CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/embdb/>).  
 CC EMBL: X06556; CAA29855.1; -  
 DR PIR: S00532; S00532  
 DR PIR: A24047; A24047  
 DR InterPro: IPR00500; Connexin.  
 DR InterPro: IPR002261; Connexin43.  
 DR Pfam: PF00029; connexin; 1.  
 DR Pfam: PF03508; Connexin43; 1.  
 DR PRINTS: P000206; CONNEXIN  
 DR SMART: SM00037; CNX; 1.  
 DR PROSITE: PS00407; CONNEXINS\_1; 1.  
 DR PROSITE: PS00408; CONNEXINS\_2; 1.  
 FT Gap junction; Transmembrane.  
 FT INIT MET 0 0  
 FT DOMAIN 1 22 CYTOPLASMIC (PROBABLE).  
 FT TRANSMEM 23 43 PROBABLE.  
 FT TRANSMEM 44 75 EXTRACELLULAR (PROBABLE).  
 FT TRANSMEM 76 96 PROBABLE.  
 FT DOMAIN 97 154 CYTOPLASMIC (PROBABLE).  
 FT TRANSMEM 155 175 PROBABLE.  
 FT TRANSMEM 176 206 EXTRACELLULAR (PROBABLE).  
 FT TRANSMEM 207 227 PROBABLE.  
 FT DOMAIN 228 381 CYTOPLASMIC (PROBABLE).  
 FT CONFLICT 1 1 G -> A (IN REF. 3).  
 FT CONFLICT 15 15 A -> T (IN REF. 2).  
 FT CONFLICT 27 27 V -> I (IN REF. 3).  
 SQ SEQUENCE 381 AA: 42900 MW: 9888907.459 PCRR CWC64;  
 Query Match 1.38; Score 7; EB 1; Length 381;  
 Best Local Similarity 100.0%; Tied. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 VNMVEMHL 509  
 DB 120 VNMVEMHL 126

RESULT 12

CC LHX9\_HUMAN STANDARD: PRT: 388 AA.  
 AC U9N059; U9N070; U9BY06;  
 DT 15-OCT-2001 (rel. 40, created)  
 DT 16-OCT-2001 (rel. 40, last sequence update)  
 DT 15-JUN-2002 (rel. 41, last annotation update)  
 DE LIM/homeobox protein Lhx9.  
 GN LHX9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Testis.  
 RC MEDLINE=21291005; PubMed=11997841.  
 RA Ottolenghi C., Moreira-Filho C., Mendonca R.R., Barbieri M.,  
 RA Feliouss M., Berkovitz G.D., McElreavey R.;  
 RT "Absence of mutations involving the LIM homeobox domain gene LHX9 in 46,XY gonadal agenesis and dysgenesis."  
 RL J. Clin. Endocrinol. Metab. 86:2465-2469(2001).  
 CC -1- FUNCTION: INVOLVED IN GONADAL DEVELOPMENT (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN PRINTS 2 CONC

DR	EMBL: AJ277916:	CAB98128.1:	ALT-SEQ.
DR	EMBL: AJ277917:	CAB98128.1:	JOINED.
DR	EMBL: AJ277918:	CAB98128.1:	JOINED.
DR	EMBL: AJ277919:	CAB98128.1:	JOINED.
DR	EMBL: AJ277920:	CAB98128.1:	JOINED.
DR	EMBL: AJ296372:	CAC33174.1:	-
DR	HSSP: P06601:	1F3L	-
DR	Genew: H0NC14222:	LHX9.	-
DR	MIM: 606066:	-	-
DR	InterPro: IPR001356:	Homeobox	-
DR	InterPro: IPR001781:	LIM.	-
DR	Pfam: PF00046:	homeobox: 1.	-
DR	Pfam: PF00412:	LIM: 2	-
DR	Proteom: P000010:	Homeobox: 1	-
DR	Proteom: P000094:	LIM: 2	-
DR	SMART: SM00189:	HOX: 1.	-
DR	SMART: SM0152:	LIM: 2	-
DR	PROSITE: PS00478:	LIM_DOMAIN_1: 2.	-
DR	PROSITE: PS00023:	LIM_DOMAIN_2: 2.	-
DR	PROSITE: PS00027:	HOXB9OX_1: 1	-
DR	PROSITE: PS00971:	HOMEBOX_2: 1.	-
DR	Homeobox: DNA binding:	Nuclear protein; Repeat; LIM domain;	-
DR	Meta-binding: zinc:	-	-
DM	DOMAIN	62	LIM 1
DM	DOMAIN	124	LIM 2.
DM	DOMAIN	258	HOMEBOX.
DM	SEQUENCE	388 AA.	ADDC6914UD7J386 CRIC4,

RESULT 13			
1D	LHX9_MOUSE	STANDARD:	PRG: 388 AA.
AC	Q9MUL2, Q9MUL4, Q9VYQ5, Q9VYQ9, Q9VYQ1.		
DT	30-MAY-2000 (Ref. 39, "Treated")		
DT	30-MAY-2000 (Ref. 39, "Last sequence update")		
DT	15-JUN-2002 (Ref. 41, "Last annotation update")		
DE	LIM/homeobox protein Lhx9.		
CN	LHX9		
OS	Mus musculus (Mouse)		
CA	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCHI_Taxid=100930;		
RN	[1]		
RP	SEQUENCE OF 1-300 FROM N.A.		
RC	STRAIN=C57BL/6;		
RA	MEDLINE=94094964; PubMed=9486506;		
RA	Relaux S., Rogard M., Hach I., Besson M.J.:		
RT	"Lhx9, a novel LIM-homeodomain gene expressed in the developing		
RT	forebrain.";		
KL	J. Neurosci., 19:783-793(1999).		
RN	[2]		
RP	SEQUENCE OF 11-388 FROM N.A.		
RC	STRAIN=NIH Swiss;		
RA	MEDLINE=99264293; PubMed=10330499;		
RA	Hotteluzzi S., Porter F.D., Pitts A., Kumar M., Agulnik A., Wassif C.,		

CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usurped or altered for commercial  
CC purposes. For more information, please contact [swissprot@ebi.ac.uk](mailto:swissprot@ebi.ac.uk)  
CC or send an email to [license@sb.slb.ch](mailto:license@sb.slb.ch).

DR	InterPro: IPR001356; Homeobox.
DR	InterPro: IPR001781; LIM.
DR	Pfam: PF00046; homeobox. 1.
DR	Pfam: PF00412; LIM. 2.
DR	Pfam: Pfam010; Homeobox. 1.
DR	ProDom: P000094; LIM. 2.
DR	SMART: SM00389; HOX. 1.
DR	SMART: SM00332; LIM. 2.
DR	ProSITE: PS00027; Homeobox 1. 1.
DR	ProSITE: PS00071; Homeobox 2. 1.
DR	ProSITE: PS00478; LIM_DOMAIN_1. 2.
DR	ProSITE: PS00023; LIM_DOMAIN_2. 2.
KW	Homeobox; DNA-binding; Nuclear protein; Repeat; LIM domain; Metal-binding; Zinc; Alternative splicing.
FT	DOMAIN 62 114
FT	DOMAIN 124 177
FT	DNA_BIND 258 317
FT	VARSPLIT 404 404
FT	HOMEBOX. VFWJAVAKRPPIILKQFNGAVKAGLILAFHSAISATLTPGATILILINPTVATVTVISMDSHEHMSQITLITNLF -> GEOILCHYQSISREKLKIP (IN ISOPFORM ALPHA).
FT	CONFLICT 49 49 A -> T (IN REF. 2).
FT	CONFLICT 153 153 S -> F (IN REF. 2).
SO	SEQUENCE 388 AA; 42986 MW; C207356A687B7E2 CRC64;
Query Match	1.3%; Score 7; DB 1; Length 388;
Host Local	100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 387 KDGSIVC 393  
 1111111  
 105 KDGSIVC 111

DB 105 KDGSIVC 111

RESULT 14  
 PCY2\_HUMAN  
 ID PCY2\_HUMAN STANDARD: PRT: 389 AA  
 AC 099447;  
 DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Ethanolamine-phosphate cytidyltransferase (EC 2.7.7.14)  
 DE (Phosphorylethanolamine transferase) (CNP:phosphoethanolamine  
 DE cytidyltransferase).  
 GN PCYT2.  
 OS Homo sapiens (Human)  
 OC Eukaryota, Metazoa; Chordata, Claviata, Vertebrata, Euteleostomi,  
 OC Mammalia, Primates; Primates, Catarrhini, Hominoidea, Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9723903; PubMed=9083101;  
 RA Nakashima A., Hosaka K., Mikawa J.;  
 RT "Cloning of a human cDNA for CNP-phosphoethanolamine  
 RT cytidyltransferase by complementation in vivo of a yeast mutant.",  
 RL J. Biol. Chem. 272:9567-9572(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases  
 CC -1- CATALYTIC ACTIVITY: GTP + ethanolamine phosphate = diphosphate +  
 CC CDP-ethanolamine.  
 CC -1- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS; CDP-ETHANOLAMINE SYNTHESIS;  
 CC SECOND STEP.  
 CC -1- TISSUE SPECIFICITY: STRONGEST EXPRESSION IN LIVER, HEART, AND  
 CC SKELETAL MUSCLE.  
 CC -1- SIMILARITY: BELONGS TO THE CYTIDYLYLTRANSFERASE FAMILY  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
 CC  
 CC FMR1: D84307; BAA1231.1;  
 DR FMR1: BC000351; AA000351.1;  
 DR Gene: HGNC:8756; PCYT2.  
 DR MIM: 602679;  
 DR InterPro: IPR004821; Cyt\_tran\_rel.  
 DR InterPro: IPR004820; Cytidylyltransf.  
 DR Pfam: PF01467; Cytidylyltransf. 2.  
 DR TIGRFAMs: TIGR00125; Cyt\_tran\_rel. 2.  
 KM Transferase, Nucleocytidyltransferase, Phospholipid biosynthe-sis  
 FT DOMAIN 20 194 CATALYTIC (POTENTIAL).  
 ST SEQUENCE 389 AA; 43835 MW; 13FEB8E87FPA7F CRC64,

Query Match 1.3%; Score 7; DR 1; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 TLETLDK 157  
 1111111  
 98 TLETLDK 104

DB 98 TLETLDK 104

RESULT 15  
 ARGI\_RACSU

11 ARGI\_RACSU STANDARD: PRT: 406 AA.  
 AC P36843;  
 DT 01-JUN-1994 (Rel. 29, created)  
 DT 01-JUN-1994 (Rel. 29, last sequence update)  
 DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Arginine biosynthesis bifunctional protein argJ [includes: glutamate  
 DE N-acetyltransferase (EC 2.3.1.35) (ornithine acetyltransferase)  
 DE (ornithine transacetylase) (OATASE); Amino-acid acetyltransferase  
 DE (EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)].  
 GN argJ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=9429722; PubMed=8025667;  
 RA O'Reilly M., Devine K.M.;  
 RT "Sequence and analysis of the citrulline biosynthetic operon argC-F  
 RT from Bacillus subtilis.";  
 RL Microbiology 143:1023-1025(1994).  
 RN [2]  
 RP SEQUENCE OF 1-80 FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98015415; PubMed=9459331;  
 RA Medina N., Vannier P., Roche R., Autret S., Lavine A., Serot S.J.;  
 RT "Sequencing of regions downstream of addA (98 degrees) and citS (289  
 RT degrees) in Bacillus subtilis.";  
 RL Microbiology 143:3305-3308(1997).  
 CC -1- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + L-glutamate = L-  
 CC ornithine + N-acetyl-L-glutamate.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + L-glutamate = CoA + N-acetyl-L-  
 CC glutamate.  
 CC -1- PATHWAY: FIRST AND FIFTH STEPS IN ARGININE BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
 CC  
 CC FMR1: Z6915; CAA81544.1;  
 DR FMR1: Z79580; CAA01843.1;  
 DR FMR1: Y04476; CAA70634.1;  
 DR FMR1: Z99109; CAA12961.1;  
 DR FMR1: Z99110; CAA12977.1;  
 DR FMR1: Z9429; S9429.  
 DR Subtilast: B01002; argJ.  
 DR InterPro: IPR002813; ArgJ.  
 DR Pfam: PF01960; ArgJ; 1.  
 DR Prodom: PD04193; ArgJ; 1.  
 DR TIGRFAMs: TIGR00120; ArgJ; 1.  
 KM Arginine biosynthesis; Transferase; Acyltransferase;  
 KM Acyltransferase; enzyme, Complete proteome;  
 ST SEQUENCE 406 AA; 44464 MW; 14C4192055D9571 CRC64;

Query Match 1.3%; Score 7; DR 1; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 VGSNLVK 376  
 1111111  
 306 VGSNLVK 312

DB 306 VGSNLVK 312

Search completed: July 15, 2003, 09:41:11  
 Job time : 26 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:38:25 ; Search time 26 seconds

(without alignments)  
595.248 Million cell updates/sec

Title: US-10-053-192-1

Perfect score: 526  
Sequence: 1 METIFNPKKEHPPIKAEV .....MHDLHGMFIPIQNDLAETTE 526

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/laa/5A-COMB-pep.\*  
2: /cgn2\_6/ptodata/1/laa/5B-COMB-pep.\*  
3: /cgn2\_6/ptodata/1/laa/6A-COMB-pep.\*  
4: /cgn2\_6/ptodata/1/laa/6B-COMB-pep.\*  
5: /cgn2\_6/ptodata/1/laa/PCITUS-COMB-pep.\*  
6: /cgn2\_6/ptodata/1/laa/backlist1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.5	205	4	US-09-385-259-3 Sequence 3, Appl
2	8	1.5	205	4	US-09-645-370-3 Sequence 3, Appl
3	8	1.5	388	1	US-08-445-640-12 Sequence 12, Appl
4	8	1.5	388	3	US-08-170-558-12 Sequence 12, Appl
5	8	1.5	388	3	US-08-447-314-12 Sequence 12, Appl
6	8	1.5	388	3	US-08-445-461-12 Sequence 12, Appl
7	8	1.5	533	1	US-08-488-305A-6 Sequence 6, Appl
8	8	1.5	533	4	US-09-385-259-2 Sequence 2, Appl
9	8	1.5	533	4	US-09-645-370-2 Sequence 2, Appl
10	8	1.5	874	2	US-08-456-647B-6 Sequence 6, Appl
11	8	1.5	874	2	US-08-237-401A-6 Sequence 6, Appl
12	8	1.5	880	1	US-08-445-640-10 Sequence 10, Appl
13	8	1.5	880	3	US-08-170-558-10 Sequence 10, Appl
14	8	1.5	880	3	US-08-447-314-10 Sequence 10, Appl
15	8	1.5	880	3	US-08-445-461-10 Sequence 10, Appl
16	8	1.5	82	2	US-08-117-952-756 Sequence 756, App
17	7	1.3	82	2	US-08-117-952-756 Sequence 756, App
18	7	1.3	183	4	US-09-134-001C-5098 Sequence 16, Appl
19	7	1.3	267	4	US-09-292-858A-15 Sequence 15, Appl
20	7	1.3	479	4	US-09-442-100-13 Sequence 13, Appl
21	7	1.3	480	4	US-08-477-928A-4 Sequence 4, Appl
22	7	1.3	699	4	US-09-457-040B-18 Sequence 18, Appl
23	7	1.3	771	4	US-09-090-793-8 Sequence 8, Appl
24	7	1.3	780	1	US-08-375-709-13 Sequence 13, Appl
25	7	1.3	780	1	US-08-752-929-13 Sequence 13, Appl
26	7	1.3	1085	1	US-08-431-080-28 Sequence 28, Appl
27	7	1.3	1085	2	US-08-938-514-28 Sequence 28, Appl

28	7	1.3	1085	4	US-09-345-294-28	Sequence 28, Appl
29	7	1.3	1142	4	US-09-106-075A-89	Sequence 89, Appl
30	7	1.3	1146	4	US-08-914-993-6	Sequence 6, Appl
31	6	1.1	15	2	US-08-937-102-20	Sequence 20, Appl
32	6	1.1	15	2	US-08-937-102-21	Sequence 21, Appl
33	6	1.1	15	2	US-08-937-102-22	Sequence 22, Appl
34	6	1.1	27	3	US-08-331-625A-18	Sequence 18, Appl
35	6	1.1	27	4	US-09-494-151-18	Sequence 18, Appl
36	6	1.1	29	2	US-08-859-201-21	Sequence 21, Appl
37	6	1.1	35	4	US-08-810-009-14	Sequence 14, Appl
38	6	1.1	41	4	US-08-469-260A-382	Sequence 382, App
39	6	1.1	52	2	US-08-466-583-7	Sequence 7, Appl
40	6	1.1	52	5	PCT-US95-07830-7	Sequence 7, Appl
41	6	1.1	54	4	US-09-187-788-49	Sequence 49, Appl
42	6	1.1	54	4	US-09-139-600-44	Sequence 44, Appl
43	6	1.1	70	2	US-08-845-256-3	Sequence 3, Appl
44	6	1.1	70	4	US-09-208-210-3	Sequence 3, Appl
45	6	1.1	89	4	US-08-936-155A-474	Sequence 474, App

## ALIGNMENTS

```

RESULT 1
US-09-385-259-3
; Sequence 3, Application US/09385259
; Patent No. 6201114
; GENERAL INFORMATION:
; APPLICANT: Aquilre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF CHROMOSOMAL STATIONARY NIGHT BLINDNESS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/385,259
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-385-259-3

Query Match      1.5%, Score 8, DB 4, Length 205;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

OY      99 PDPCKNIF 106
      |||||||
DB      109 PDPCKNIF 116

RESULT 2
US-09-645-370-3
; Sequence 3, Application US/09645370
; Patent No. 6428958
; GENERAL INFORMATION:
; APPLICANT: Aquilre, Gustavo D.
; APPLICANT: Acland, Gregory M.
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS
; FILE REFERENCE: 19603/2481
; CURRENT APPLICATION NUMBER: US/09/645,370
; EARLIER FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/385,259
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: 60/103,219
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 8

```

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 3  
LENGTH: 205  
TYPE: PRT  
ORGANISM: Canis familiaris  
US-09-645-370-3

Query Match 1.5%: Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDPCKNIF 106  
Db 109 PDPCKNIF 116

RESULT 3  
US-08-445-640-12  
Sequence 12, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445.640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 188 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-640-12

Query Match 1.5%: Score 8; DB 1; Length 388;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGIALLHS 59  
Db 210 DGIALLHS 217

RESULT 4  
US-08-170-558-12

Sequence 12, Application US/08170558  
Patent No. 6001621  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170.558  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-170-558-12

Query Match 1.5%: Score 8; DB 3; Length 388;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGIALLHS 59  
Db 210 DGIALLHS 217

RESULT 5  
US-08-447-314-12  
Sequence 12, Application US/08447314  
Patent No. 6087144  
GENERAL INFORMATION:  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genetech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,314  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-447-314-12

Query Match 1.5%, Score 8, DB 3, Length 388,  
Best Local Similarity 100.0%, Pred. No. 13,  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
DB 210 DGLALLHS 217

RESULT 6  
US-08-445-461-12  
Sequence 12, Application US/08445461  
Patent No. 6096527  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genetech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,461  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-461-12

Query Match 1.5%, Score 8, DB 3, Length 388;  
Best Local Similarity 100.0%, Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
DB 210 DGLALLHS 217

RESULT 7  
US-08-488-305A-6  
Sequence 6, Application US/08488305A  
Patent No. 5679772  
GENERAL INFORMATION:  
APPLICANT: Birk, Claes Olof; Eriksson, Ulf; Peterson, Per A.  
TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which  
Bind Thereeto, Nucleic Acid Sequence Coding  
Patent No. 5679772  
TITLE OF INVENTION: Therefore, And Uses Thereof  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Feltz & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,305A  
FILING DATE: 7-JUNE-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohli, Vinet  
REGISTRATION NUMBER: 37,003  
REFERENCE/DOCKET NUMBER: LUD 5280.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9300  
TELEFAX: (212) 688-3884  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 533 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-488-305A-6

Query Match 1.5%, Score 8, DB 1, Length 533;  
Best Local Similarity 100.0%, Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PDPCKNIF 106  
DB 109 PDPCKNIF 116

RESULT 8  
US-09-385-259 2  
Sequence 2, Application US/09385259  
Patent No. 620114  
GENERAL INFORMATION:  
APPLICANT: Aquilite, Gustavo D.  
APPLICANT: Aquilite, Gregory M.  
APPLICANT: Ray, Kunal  
TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS  
TITLE OF INVENTION: IN DOGS  
FILE REFERENCE: 19603/2481  
CURRENT APPLICATION NUMBER: US/09/385,259  
CURRENT FILING DATE: 1999-08-30  
EARLIER APPLICATION NUMBER: 60/103,219  
EARLIER FILING DATE: 1998-10-06  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 543  
TYPE: PRT  
ORGANISM: Canis familiaris  
US-09-385-259 2

Query Match 1.5% Score 8; DB 4; Length 533;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDPCKNIF 106  
|||||||  
DB 109 PDPCKNIF 116

RESULT 9  
US-09-645-370-2  
Sequence 2, Application US/09645370  
Patent No. 6428958  
GENERAL INFORMATION:  
APPLICANT: Aquilite, Gustavo D.  
APPLICANT: Acland, Gregory M.  
APPLICANT: Ray, Kunal  
TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS  
TITLE OF INVENTION: IN DOGS  
FILE REFERENCE: 19603/2481  
CURRENT APPLICATION NUMBER: US/09/645,370  
CURRENT FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: 09/385,259  
PRIOR FILING DATE: 1999-08-30  
PRIOR APPLICATION NUMBER: 60/103,219  
PRIOR FILING DATE: 1998-10-06  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 533  
TYPE: PRT  
ORGANISM: Canis familiaris  
US-09-645-370-2

Query Match 1.5% Score 8; DB 4; Length 533;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 PDPCKNIF 106  
|||||||  
DB 109 PDPCKNIF 116

RESULT 10  
US-08-456-647B-6  
Sequence 6, Application US/08456647B  
Patent No. 5811516  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.

TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,647B  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/237,401  
FILING DATE: 02-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/864,406  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell Ph.D., John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 07251/007002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 874 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-456-647B-6

Query Match 1.5% Score 8; DB 2; Length 874;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 DGLALHS 59  
|||||||  
DB 234 DGLALHS 241

RESULT 11  
US-08-237-401A-6  
Sequence 6, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:



APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO.: 6  
SEQUENCE CHARACTERISTICS:  
LENGTH: 874 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-6

Query Match 1.5%; Score 8; DB 2; Length 874;  
Best Local Similarity 100.0%; Pred No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
|||||  
DB 234 DGLALLHS 241

RESULT 12  
US-08-445-640-10  
Sequence 10, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie P.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Barton, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: palin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO.: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 880 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-445-640-10  
Query Match 1.5%; Score 8; DB 1; Length 880;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
|||||  
DB 240 DGLALLHS 247

RESULT 13  
US-08-170-558-10  
Sequence 10, Application US/08170558  
Patent No. 6001621  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Barton, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 460 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: palin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170,558  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO.: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 880 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-170-558-10

Query Match 1.5%; Score 8; DB 3; Length 880;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
|||||  
DB 240 DGLALLHS 247

RESULT 14  
US-08-447-314-10  
Sequence 10, Application US/08447314  
Patent No. 6087144  
GENERAL INFORMATION:  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.

```

? APPLICANT: Baron, Will F.
? TITLE OF INVENTION: Protein Tyrosine Kinases
? NUMBER OF SEQUENCES: 35
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 1RM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: patin (Genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/447,314
? FILING DATE: 22-MAY-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/170558
? FILING DATE: 20-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/157563
? FILING DATE: 23-NOV-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 854C1D2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 880 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
?
? US-08-447-314-10
?
? Query Match 1.5%; Score 8; DB 3; Length 880;
? Best Local Similarity 100.0%; Pred. No. 27;
? Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 52 DGLALLHS 59
? 11111111
? Db 240 DGLALLHS 247

```

```

? RESULT 15
? US-08-445-461-10
? Sequence 10, Application US/08445461
? Patent No. 6096527
? GENERAL INFORMATION:
? APPLICANT: Godowski, Paul J.
? APPLICANT: Mark, Melanie R.
? APPLICANT: Scadden, David T.
? APPLICANT: Baker, Kevin P.
? APPLICANT: Baron, Will F.
? TITLE OF INVENTION: Protein Tyrosine Kinases
? NUMBER OF SEQUENCES: 35
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5 25 inch, 360 kb floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: patin (Genentech)

```

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/445,461
? FILING DATE: 22-MAY-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/170558
? FILING DATE: 20-DEC-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/157563
? FILING DATE: 23-NOV-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Hasak, Janet E.
? REGISTRATION NUMBER: 28,616
? REFERENCE/DOCKET NUMBER: 854C3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 880 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
?
? US-08-445-461-10
?
? Query Match 1.5%; Score 8; DB 3; Length 880;
? Best Local Similarity 100.0%; Pred. No. 27;
? Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 52 DGLALLHS 59
? 11111111
? Db 240 DGLALLHS 247

```

Search completed: July 15, 2003, 09:43:17  
 Job time : 28 secs

Genome version 5.1  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2003, 09:41:16 : Search time: 56 seconds  
(without alignments)  
1113.396 Million cells updated/seen

Title: US-10-053-192-1

526

Sequence: 1 METFNPKKEHPPTKAEV ..... MHTDHPMTDPNDICAFTR 526

Scoring table: Gapop 60 0, Gapext 60 0

Searched: 445758 seqs, 116419773 residues

Word size: 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Published\_Applications\_AA\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB\_PEP.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB\_PEP.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*

Prod. No. is the number of results produced by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	526	100	6	US-10-053-192-1	Sequence 1, Appl 1
2	506	96.2	506	US-10-053-192-4	Sequence 4, Appl 1
3	18	3.4	18	US-10-053-192-6	Sequence 6, Appl 1
4	18	3.4	18	US-10-053-192-7	Sequence 7, Appl 1
5	8	1.5	8	US-10-053-192-3	Sequence 3, Appl 1
6	8	1.5	386	US-09-985-675-5	Sequence 5, Appl 1
7	8	1.5	388	US-09-223-428-12	Sequence 12, Appl 1
8	8	1.5	529	US-10-053-192-5	Sequence 5, Appl 1
9	8	1.5	580	US-09-928-457-37	Sequence 37, Appl 1
10	8	1.5	850	US-09-985-675-2	Sequence 2, Appl 1
11	8	1.5	874	US-09-158-722-6	Sequence 6, Appl 1
12	8	1.5	874	US-09-985-675-1	Sequence 1, Appl 1
13	8	1.5	880	US-09-223-428-10	Sequence 10, Appl 1
14	7	1.3	23	US-09-809-391-621	Sequence 621, Appl 1
15	7	1.3	32	US-09-804-761-34560	Sequence 34560, Appl 1
16	7	1.3	53	US-09-789-561-118	Sequence 118, Appl 1
17	7	1.3	53	US-09-789-561-149	Sequence 149, Appl 1
18	7	1.3	53	US-09-789-561-149	Sequence 149, Appl 1
19	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1

20	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
21	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
22	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
23	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
24	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
25	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
26	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
27	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
28	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
29	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
30	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
31	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
32	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
33	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
34	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
35	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
36	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
37	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
38	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
39	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
40	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
41	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
42	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
43	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
44	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1
45	7	1.3	108	US-09-764-891-5170	Sequence 5170, Appl 1

## ALIGNMENTS

RESULT 1	US-10-053-192-1	Sequence 1, Appl 1	US-10-053-192-1	Sequence 1, Appl 1
1	100	100	100	100
2	100	100	100	100
3	100	100	100	100
4	100	100	100	100
5	100	100	100	100
6	100	100	100	100
7	100	100	100	100
8	100	100	100	100
9	100	100	100	100
10	100	100	100	100
11	100	100	100	100
12	100	100	100	100
13	100	100	100	100
14	100	100	100	100
15	100	100	100	100
16	100	100	100	100
17	100	100	100	100
18	100	100	100	100
19	100	100	100	100
20	100	100	100	100
21	100	100	100	100
22	100	100	100	100
23	100	100	100	100
24	100	100	100	100
25	100	100	100	100
26	100	100	100	100
27	100	100	100	100
28	100	100	100	100
29	100	100	100	100
30	100	100	100	100
31	100	100	100	100
32	100	100	100	100
33	100	100	100	100
34	100	100	100	100
35	100	100	100	100
36	100	100	100	100
37	100	100	100	100
38	100	100	100	100
39	100	100	100	100
40	100	100	100	100
41	100	100	100	100
42	100	100	100	100
43	100	100	100	100
44	100	100	100	100
45	100	100	100	100

QY 181 ILNMGTSIVDKGRKYVLEKIPSSVPEKEKKKSCFKHLEVCSIPSRSLDPSYHSGI 240  
DB 181 ILNMGTSIVDKGRKYVLEKIPSSVPEKEKKKSCFKHLEVCSIPSRSLDPSYHSGI 240  
QY 241 TENYIVLEIOPFKLDIYKLTATAYIRGVNMASSCLSPFKEDKTMWFEPRDKTKKEVSTKFT 300  
DB 241 TENYIVLEIOPFKLDIYKLTATAYIRGVNMASSCLSPFKEDKTMWFEPRDKTKKEVSTKFT 300  
QY 301 DALVLYHHINAYEEDGHVFDIYARNDSLYDMFYLLKLDKDFEVNNKLTSLPTCKRFVY 360  
DB 301 DALVLYHHINAYEEDGHVFDIYARNDSLYDMFYLLKLDKDFEVNNKLTSLPTCKRFVY 360  
QY 361 PLOYDKDAEVSNNLVKLPSTATAVKEKDSICYOPPELCEGIELPRVNDYNGKRYKYV 420  
DB 361 PLOYDKDAEVSNNLVKLPSTATAVKEKDSICYOPPELCEGIELPRVNDYNGKRYKYV 420  
QY 421 ATEVQMSPVPTKIAKLNVQTKEVYLMHGEDHCWSEPIFVSPDAREDEGVVLYCVVSE 480  
DB 421 ATEVQMSPVPTKIAKLNVQTKEVYLMHGEDHCWSEPIFVSPDAREDEGVVLYCVVSE 480  
QY 481 PNKAPFLILIDAKTFKELGRATVNVEMHLLHGMFIPONDLAGETE 526  
DB 481 PNKAPFLILIDAKTFKELGRATVNVEMHLLHGMFIPONDLAGETE 526

RESULT 2  
US-10-053-192-4  
Sequence 4, Application US/10053192  
Publication No. US20030087336A1

GENERAL INFORMATION:  
APPLICANT: BACHMANN, Heinrich  
APPLICANT: BRUGGER, Roland  
APPLICANT: FRIEDELIN, Arno M  
APPLICANT: WIRTZ, Gabriele M  
APPLICANT: WOGGON, Wolf-Dietrich  
APPLICANT: WYSS, Adrian  
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...  
CURRENT APPLICATION NUMBER: 08/10/053,192  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 103382.0  
PRIOR FILING DATE: 1999-02-22  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 4  
LENGTH: 506  
TYPE: PRT  
ORGANISM: CHICKEN  
US-10-053-192-4

Query Match 96.2%; Score 506; DB 9, Length 506;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 10 EEHPEPIKAEVQGLPPIWIGVLLRNGPCMHITIGTKYHMFPGALLHSTFFKNGEYVY 69  
DB 1 EEHPEPIKAEVQGLPPIWIGVLLRNGPCMHITIGTKYHMFPGALLHSTFFKNGEYVY 69  
QY 70 RSKILRSIDYNCNIEANRIIVSEFGIMAYPDPCKNIFAKAFYSLSHTIPEFTDNLINIM 129  
DB 61 RSKILRSIDYNCNIEANRIIVSEFGIMAYPDPCKNIFAKAFYSLSHTIPEFTDNLINIM 120  
QY 140 KTGDYVATSETNFIRKIDPQLETLDDKYVSKYAVANLATSHPHYDSAGNILLNGSTIV 189  
DB 121 KTGDYVATSETNFIRKIDPQLETLDDKYVSKYAVANLATSHPHYDSAGNILLNGSTIV 180  
QY 190 DKGCTKYVLFKIPSSVPEKEKKKSCFKHLEVCSIPSRSLDPSYHSGIENYIVLE 249  
DB 181 DKGCTKYVLFKIPSSVPEKEKKKSCFKHLEVCSIPSRSLDPSYHSGIENYIVLE 240

QY 250 QPEKLDIVKLTATAYIRGVNMASSCLSPFKEDKTMWFEPRDKTKKEVSTKRYTIALVYHHI 409  
DB 241 QPEKLDIVKLTATAYIRGVNMASSCLSPFKEDKTMWFEPRDKTKKEVSTKRYTIALVYHHI 400  
QY 310 NAYEEDGHVFDIYARNDSLYDMFYLLKLDKDFEVNNKLTSLPTCKRFVNDYNGKRYKYV 469  
DB 301 NAYEEDGHVFDIYARNDSLYDMFYLLKLDKDFEVNNKLTSLPTCKRFVNDYNGKRYKYV 460  
QY 370 VGSNLVYLPTSATAVKEKDSICYOPPELCEGIELPRVNDYNGKRYKYVATEVQMSVY 429  
DB 361 VGSNLVYLPTSATAVKEKDSICYOPPELCEGIELPRVNDYNGKRYKYVATEVQMSVY 420  
QY 430 PTKIAKLNVQTKEVYLMHGEDHCWSEPIFVSPDAREDEGVVLYCVVSEPNKAPFLIL 489  
DB 421 PTKIAKLNVQTKEVYLMHGEDHCWSEPIFVSPDAREDEGVVLYCVVSEPNKAPFLIL 480  
QY 490 LDKTFKELGRATVNVEMHLLHGMF 519  
DB 481 LDKTFKELGRATVNVEMHLLHGMF 506

RESULT 3  
US-10-053-192-6  
Sequence 6, Application US/10053192  
Publication No. US20030087336A1

GENERAL INFORMATION:  
APPLICANT: BACHMANN, Heinrich  
APPLICANT: BRUGGER, Roland  
APPLICANT: FRIEDELIN, Arno M  
APPLICANT: WIRTZ, Gabriele M  
APPLICANT: WOGGON, Wolf-Dietrich  
APPLICANT: WYSS, Adrian  
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...  
CURRENT APPLICATION NUMBER: 08/10/053,192  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 103382.0  
PRIOR FILING DATE: 1999-02-22  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 6  
LENGTH: 18  
TYPE: PRT  
ORGANISM: CHICKEN  
US-10-053-192-6

Query Match 3.4%; Score 18; DB 9, Length 18;  
Best Local Similarity 100.0%; Pred. No. 4, 6e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NKEHPEPIKAEVQGLP 25  
DB 1 NKEHPEPIKAEVQGLP 18  
RESULT 4  
US-10-053-192-7  
Sequence 7, Application US/10053192  
Publication No. US20030087336A1  
GENERAL INFORMATION:  
APPLICANT: BACHMANN, Heinrich  
APPLICANT: BRUGGER, Roland  
APPLICANT: FRIEDELIN, Arno M  
APPLICANT: WIRTZ, Gabriele M  
APPLICANT: WOGGON, Wolf-Dietrich  
APPLICANT: WYSS, Adrian  
TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES, ...  
CURRENT APPLICATION NUMBER: 08/10/053,192

1 CURRENT FILING DATE: 2002-01-15  
2 PRIOR APPLICATION NUMBER: 103382.0  
3 PRIOR FILING DATE: 1999-02-22  
4 NUMBER OF SEQ ID NOS: 10  
5 SOFTWARE: Patent In Ver. 2.1  
6 SEQ ID NO: 7  
7 LENGTH: 18  
8 TYPE: PPT  
9 ORGANISM: CHICKEN  
10 US-10-053-192-7

Query Match 3.4% Score 18; DB 9; Length 18;  
Best Local Similarity 100.0%; Pred No. 3; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NKEHEPEIKAEVQGLP 25  
1 NKEHEPEIKAEVQGLP 18

Db 1 NKEHEPEIKAEVQGLP 18

RESULT 5  
US-10-053-192-3  
1 Sequence 3, Application US/10053192  
2 Publication No. US20030087336A1  
3 GENERAL INFORMATION:  
4 APPLICANT: BACHMANN, Heinrich  
5 APPLICANT: BRUGER, Roland M  
6 APPLICANT: FRIEDLEIN, Arno M  
7 APPLICANT: WIRTS, Gabriele M  
8 APPLICANT: WOGGIN, Wolf-Dietrich  
9 APPLICANT: WYSS, Adrian  
10 APPLICANT: WYSS, Markus  
11 TITLE OF INVENTION: BETA-BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
12 TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE  
13 FILE REFERENCE: B, B-CAROTENE 15,15'-DIOXYGENASES,....  
14 CURRENT APPLICATION NUMBER: US/10/053,192  
15 PRIOR FILING DATE: 2002-01-15  
16 PRIOR APPLICATION NUMBER: 103382.0  
17 PRIOR FILING DATE: 1999-02-22  
18 NUMBER OF SEQ ID NOS: 10  
19 SOFTWARE: Patent In Ver. 2.1  
20 SEQ ID NO: 3  
21 LENGTH: 8  
22 TYPE: PPT  
23 ORGANISM: CHICKEN  
24 US-10-053-192-3

Query Match 1.5% Score 8; DB 9; Length 8;  
Best Local Similarity 100.0%; Pred No. 4e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AEVQGLP 25  
1 AEVQGLP 8

Db 1 AEVQGLP 8

RESULT 6  
US-09-985-675-5  
1 Sequence 5, Application US/09985675  
2 Patent No. US20020137168A1  
3 GENERAL INFORMATION:  
4 APPLICANT: GROSIER, PHILIP S.  
5 APPLICANT: GROSIER, KATHRYN F  
6 TITLE OF INVENTION: DEVELOPMENTAL TYROSINE KINASES AND  
7 THEIR LIGANDS  
8 NUMBER OF SEQUENCES: 16  
9 CORRESPONDENCE ADDRESS:  
10 ADDRESSEE: NIXON & VANDERHYTE P.C.  
11 STREET: 1100 NORTH GLENN ROAD  
12 CITY: ARLINGTON  
13 STATE: VIRGINIA  
14 COUNTRY: U S A  
15 ZIP: 22201-4714

1 COMPUTER READABLE FORM:  
2 MEDICIN TYPE: Floppy disk  
3 COMPUTER: IBM PC compatible  
4 OPERATING SYSTEM: PC-TOS/MS-TOS  
5 SOFTWARE: Patent In Ver. 2.1, Version #1.25  
6 CURRENT APPLICATION DATA:  
7 APPLICATION NUMBER: US/09/985,675  
8 FILING DATE: 05-NOV-98  
9 CLASSIFICATION: <unknown>  
10 PRIOR APPLICATION DATA:  
11 APPLICATION NUMBER: US 08/505,241  
12 FILING DATE: 15-AUG-1995  
13 ATTORNEY/AGENT INFORMATION:  
14 NAME: MITCHELL, LEONARD C.  
15 REGISTRATION NUMBER: 29,009  
16 REFERENCE/POCKET NUMBER: 175-19  
17 TELECOMMUNICATION INFORMATION:  
18 TELEPHONE: (703) 816-4000  
19 TELEFAX: (703) 816-4100  
20 INFORMATION FOR SEQ ID NO: 5:  
21 SEQUENCE CHARACTERISTICS:  
22 LENGTH: 386 AMINO ACIDS  
23 TYPE: AMINO ACID  
24 TOPOLOGY: LINEAR  
25 MOLECULE TYPE: PROTEIN  
26 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
27 US-09-985-675-5

Query Match 1.5% Score 8; DB 10; Length 386;  
Best Local Similarity 100.0%; Pred No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALHS 59  
1 DGLALHS 217

Db 210 DGLALHS 217

RESULT 7  
US-09-923-490-12  
1 Sequence 12, Application US/0923490  
2 Patent No. US20020147325A1  
3 GENERAL INFORMATION:  
4 APPLICANT: Godowski, Paul J.  
5 APPLICANT: Mark, Melanie R.  
6 APPLICANT: Scadden, David T.  
7 APPLICANT: Baker, Kevin P.  
8 APPLICANT: Batton, Will F.  
9 TITLE OF INVENTION: Protein Tyrosine Kinases  
10 NUMBER OF SEQUENCES: 35  
11 CORRESPONDENCE ADDRESS:  
12 ADDRESSEE: Genentech, Inc.  
13 STREET: 460 Point San Bruno Blvd  
14 CITY: South San Francisco  
15 STATE: California  
16 COUNTRY: USA  
17 ZIP: 94080  
18 COMPUTER READABLE FORM:  
19 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
20 COMPUTER: IBM PC compatible  
21 OPERATING SYSTEM: PC-TOS/MS-TOS  
22 SOFTWARE: Pat In (Genentech)  
23 CURRENT APPLICATION DATA:  
24 APPLICATION NUMBER: US/09/223,490  
25 FILING DATE:  
26 CLASSIFICATION:  
27 PRIOR APPLICATION DATA:  
28 APPLICATION NUMBER: 08/170,558  
29 FILING DATE:  
30 ATTORNEY/AGENT INFORMATION:  
31 NAME: Hasak, Janet E.  
32 REGISTRATION NUMBER: 28,616  
33 REFERENCE/POCKET NUMBER: 854C1  
34 TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 488 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-223-490-12

Query Match 1.5%: Score 8; DB 10; Length 388;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALHS 59  
|||||  
DB 210 DGLALHS 217

RESULT 8  
US-10-053-192-5  
Sequence 5, Application US/10053192  
Publication No. US20030087336A1  
GENERAL INFORMATION:  
APPLICANT: BACHMANN, Heinrich  
APPLICANT: BROGGER, Roland  
APPLICANT: FRIEDLEIN, Arno M  
APPLICANT: WIRTZ, Gabriele M  
APPLICANT: WOGGON, Wolf-Dietrich  
APPLICANT: WYSS, Adrian  
TITLE OF INVENTION: BETA-HETA-CAPROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE  
FILE REFERENCE: R, B-CAPROTENE 15,15'-DIOXYGENASES, ...  
CURRENT APPLICATION NUMBER: US/10/053,192  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 103382.0  
PRIOR FILING DATE: 1999-02-22  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 529  
TYPE: PRT  
ORGANISM: ROVINE  
US-10-053-192-5

Query Match 1.5%: Score 8; DB 9; Length 529;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PDPCKNF 106  
|||||  
DB 94 PDPCKNF 101

RESULT 9  
US-09-928-457-37  
Sequence 37, Application US/09928457  
Patent No. US20020164603A1  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DNA, specific proteins and peptides  
TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
TITLE OF INVENTION: for obtaining them and their biological application.  
NUMBER OF SEQUENCES: 99  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/928,457  
FILING DATE: 2001-08-14

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/214,759  
FILING DATE: 199-12-10  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 580 acids amin.s  
TYPE: acid amin.  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..580  
US-09-928-457-37

Query Match 1.5%: Score 8; DB 9; Length 580;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 VNYDNGK 414  
|||||  
DB 325 VNYDNGK 332

RESULT 10  
US-09 985 675-2  
Sequence 2, Application US/09985675  
Patent No. US20020147168A1  
GENERAL INFORMATION:  
APPLICANT: CROSTER, PHILIP S.  
TITLE OF INVENTION: DEVELOPMENTAL TYPE-SINE KINASES AND  
THEIR LIGANDS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHUYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/985,675  
FILING DATE: 05-NO. US20020147168A1 2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/505,241  
FILING DATE: 16-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MITCHARD, LEONARD C.  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 175-19  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816 4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 850 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SPQ ID NO: 2;  
US-09-985-675-2

Query Match 1.5%: Score 8; DB 10; Length 850;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
DB 210 DGLALLHS 217

## RESULT 11

US-09-158-722-6  
Sequence 6, Application US/09158722  
Publication No. US2004001348A1  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 425 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/158,722  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/456,647  
FILING DATE: 02-JUN-1995  
APPLICATION NUMBER: US 08/237,401  
FILING DATE: 02-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell Ph.D., John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 07/251/007902  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 874 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-158-722-6

Query Match 1.58, Score 8, DB 3, Length 874,  
Best Local Similarity 100.0%, Pred. No. 58;  
Matches 8; Conservative 0, Mismatches 0, Indels 0, Gaps 0,

QY 52 DGLALLHS 59  
DB 234 DGLALLHS 241

## RESULT 12

US-09-985-675-1  
Sequence 1, Application US/09985675  
Patent No. US20020137168A1  
GENERAL INFORMATION:  
APPLICANT: CROSIER, PHILIP S.  
CROSIER, KATHRYN E.  
TITLE OF INVENTION: DEVELOPMENTAL TYROSINE KINASES AND  
THEIR LIGANDS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYTE P.C.

STREET: 1100 NORTH GLENN ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/985,675  
FILING DATE: 05-NOV-2002/013/168A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/555,241  
FILING DATE: 16-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MITCHELL, LEONARD C.  
REGISTRATION NUMBER: 39,009  
REFERENCE/DOCKET NUMBER: 175-19  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 874 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-985-675-1

Query Match 1.58, Score 8, DB 10, Length 874;  
Best Local Similarity 100.0%, Pred. No. 58;  
Matches 8; Conservative 0, Mismatches 0, Indels 0, Gaps 0,

QY 52 DGLALLHS 59  
DB 234 DGLALLHS 241

## RESULT 13

US-09-223-490-10  
Sequence 10, Application US/09223490  
Patent No. US20020147325A1  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul T.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Spadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Barron, Will P.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/223,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/1/0,558  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28 616  
REFERENCE/DOCKET NUMBER: 854C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/771-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 880 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-223-490-10

Query Match: 1.58; Score 8; DB 10; Length 880;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DGLALLHS 59  
DB 240 DGLALLHS 247

RESULT 14  
US-09-809-391-621  
Sequence 621, Application US/09809391  
Publication No. US20030049618A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P2  
CURRENT APPLICATION NUMBER: US/09/809,391  
CURRENT FILING DATE: 2001-03-16  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 761  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 621  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-809-391-621

Query Match: 1.38; Score 7; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 KIPSSVP 206  
DB 12 KIPSSVP 18

RESULT 15  
US-09-864-761-34560  
Sequence 34560, Application US/09864761  
Patent No. US20020348763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Acomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/808,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34560  
LENGTH: 32  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC0034814.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.8  
OTHER INFORMATION: EXPRESSED IN H9100, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN HONE MARROW, SIGNAL = 5.1  
OTHER INFORMATION: EST HUMAN HIT: AJ129419.1, EVALUATE 2.00e-10  
OTHER INFORMATION: SWISSPROT HIT: P70501, EVALUATE 2.90e-02  
US-09-864-761-34560

Query Match: 1.38; Score 7; DB 10; Length 32;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 VAVNLAT 170  
DB 23 VAVNLAT 29

Search completed: July 15, 2003, 09:50:21  
Total time: 56 secs



GenCode version 5.1.6  
Copyright (c) 1994 - 2003 Wapgen Ltd.

## OM protein - protein search, using sw model

Run on: July 15, 2003, 09:19:49 ; Search time 46 seconds  
(without alignments)  
1275 000 Million cell updates/sec

Title: US-10-053-192-1

Perfect score: 2825

Sequence: 1 MFTFNPKRKRHRPKRPAKV.....MHLHGMELPNUNLSAETE 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671500 seqs, 200047115 residues

Total number of hits satisfying chosen parameters: 67160

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: SP archaea:\*

2: SP bacteria:\*

3: SP fungi:\*

4: SP human:\*

5: SP invertebrate:\*

6: SP mammal:\*

7: SP mhc:\*

8: SP organelle:\*

9: SP phage:\*

10: SP plant:\*

11: SP rodent:\*

12: SP virus:\*

13: SP vertebrate:\*

14: SP unclassified:\*

15: SP virois:\*

16: SP bacterioph:\*

17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

## SUMMARY

Result No.	Score	Query Match	Length	FR	IF	Description
1	2825	100.0	526	1	0.91993	Q91993 gallus gall
2	2134	75.2	566	11	0.91186	Q91186 mus muscul
3	2131	75.1	566	11	0.90899	Q90899 mus muscul
4	2118	75.0	566	11	0.91185	Q91185 ratu
5	2114	74.8	547	4	0.91416	Q91416 homo sapien
6	2107	74.6	547	4	0.91416	Q91416 homo sapien
7	1715	60.7	516	13	0.90014	Q90014 brachydanio
8	1142	40.4	549	13	0.90014	Q90014 brachydanio
9	1115	39.5	545	4	0.91008	Q91008 homo sapien
10	1112	39.4	532	11	0.91416	Q91416 mus muscul
11	1102	39.0	539	4	0.91416	Q91416 mus muscul
12	1102	39.0	556	4	0.91416	Q91416 mus muscul
13	1077	38.1	533	11	0.91205	Q91205 mus muscul
14	1075	38.1	533	4	0.91518	Q91518 mus muscul
15	1072	37.5	533	11	0.91518	Q91518 mus muscul
16	1071	37.5	533	4	0.91518	Q91518 mus muscul

17	1070	37.9	533	6	0.91771	Q91771 ceropithe
18	1070	37.9	533	6	0.91771	Q91771 ceropithe
19	1069	37.8	533	13	0.91771	Q91771 ceropithe
20	1068	37.8	533	6	0.91771	Q91771 ceropithe
21	1068	37.8	533	6	0.91771	Q91771 ceropithe
22	1064	37.7	533	13	0.91771	Q91771 ceropithe
23	993.5	35.2	483	11	0.91771	Q91771 ceropithe
24	840.5	29.4	620	5	0.91771	Q91771 ceropithe
25	838.5	29.7	620	5	0.91771	Q91771 ceropithe
26	813.5	28.8	544	5	0.91771	Q91771 ceropithe
27	778.5	27.6	556	5	0.91771	Q91771 ceropithe
28	652.5	23.1	529	11	0.91771	Q91771 ceropithe
29	541	19.2	570	10	0.91771	Q91771 ceropithe
30	445	15.8	552	10	0.91771	Q91771 ceropithe
31	437	15.5	616	10	0.91771	Q91771 ceropithe
32	395	14.0	490	16	0.91771	Q91771 ceropithe
33	376.5	13.3	497	16	0.91771	Q91771 ceropithe
34	312	11.0	442	2	0.91771	Q91771 ceropithe
35	311	11.0	538	10	0.91771	Q91771 ceropithe
36	310	11.0	538	10	0.91771	Q91771 ceropithe
37	401.5	10.7	542	10	0.91771	Q91771 ceropithe
38	280	9.9	480	16	0.91771	Q91771 ceropithe
39	279	9.9	475	16	0.91771	Q91771 ceropithe
40	276	9.8	605	10	0.91771	Q91771 ceropithe
41	269	9.5	790	5	0.91771	Q91771 ceropithe
42	259	9.5	842	5	0.91771	Q91771 ceropithe
43	254.5	9.4	504	16	0.91771	Q91771 ceropithe
44	254.5	9.4	472	16	0.91771	Q91771 ceropithe
45	264	9.3	604	10	0.91771	Q91771 ceropithe

## ALIGNMENTS

RESULT 1	ID	Q91993	PRELIMINARY:	PRT:	526 AA.
AC	Q91993				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 15, Last annotation update)			
DE	Beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).				
GN	BDOO.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SHOWN FROM N.A.				
RP	STRAIN=US-10-053-192-1; TISSUE=embryo;				
FX	MDL=20261261; PubMed=10792977;				
PA	Wysse A., Wirtz G.M., Woggon W.D., Brugger P., Wyss M., Friedlein A.,				
RA	Bachmann H., Hunziker W.;				
RT	Cloning and expression of beta,beta-carotene-15,15'-dioxygenase.*;				
PL	biochem. biophys. Res. Commun. 271:344-356(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
FC	STRAIN=US-10-053-192-1; TISSUE=embryo;				
FX	MDL=20261261; PubMed=10792977;				
PA	Wysse A., Wirtz G.M., Woggon W.D., Brugger P., Wyss M., Friedlein A.,				
RA	Bachmann H., Hunziker W.;				
RT	Cloning and expression of beta,beta-carotene-15,15'-dioxygenase.*;				
PL	biochem. biophys. Res. Commun. 271:344-356(2000).				
DP	EMBL: A1277386; C989825.1; -;				
DE	Interf. (P004434; P0655; P0655;				
DE	Plasm. P03055; P0655; P0655;				
KW	Dioxygenase; Oxidoreductase;				
SU	SEQUENCE 526 AA; 60384 MW; 8206EAAAP75P445h (C9844;				
Query Match	100.0%	Score	2825	DB	100.0%
Post-local Statistics	100.0%	Prod	No.	1	2023

Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 Db 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 QY 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 Db 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 QY 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 Db 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 QY 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 Db 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 QY 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 Db 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 QY 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 Db 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 QY 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 Db 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 QY 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 Db 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 QY 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526  
 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526  
 Db 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526

RESULT 2  
 Q9JJS6 PRELIMINARY; PRT; 566 AA.

AC Q9JJS6  
 ID Q9JJS6  
 DT 01-OCT-2000 (TREMblrel. 15, created)  
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)  
 DE beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).  
 GN BC001 OR BC001 OR BC001  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RA Weiss A., Wirtz G.M., Wogden W.D., Brugger R., Weiss M., Fildeslein A.,  
 Bachmann H., Hunziker W.;  
 RT "Expression pattern and localization of beta, beta carotene 15,15'-  
 dioxygenase in different tissues."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBPRT databases  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Regmond T.M., Gentleman S., Duncan T., Yu S., Wiggert B., Gantt F.,  
 Cunningham F.X.;  
 RT "Identification, expression, and substrate specificity of a mammalian  
 beta-carotene 15,15'-dioxygenase."  
 RL J Biol. Chem. 0-0-0(2001)  
 DB EMBL: AJ278064; CAB92531.2;  
 DB EMBL: AF271398; AAC33982.1;  
 DB MGP: MG11925923; Bcd01.  
 DB InterPro: IPR004294; PF0655.  
 DB Pfam: PF03055; RPF65; 1.

KM Dioxxygenase; Oxidoreductase;  
 SQ SEQUENCE 566 AA; 64864 MW; 1B4767815247A8D2 676764;

Query Match 75.28; Score 2124; Db 11; Length 566;  
 Best Local Similarity 71.28; Pval. No. 9.4e-150;  
 Matches 374; Conservative 89; Mismatches 64; Indels 4; Gaps 4;

QY 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 Db 1 METIFNNKKEHPPIKAEVGGGLPTWLGCVLLRNQPGHMTIGDTYNNHFPDLALHSF 60  
 QY 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 Db 61 TERNVEYYSKYLPSITYNWNIENAVIVSEFGTMAVPPCKNIFAKAFSYLSHTIPFF 120  
 QY 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 Db 121 TDNCLINIMKTGGDYATSETNFIKIDPQTLTDLKVDYSKYVAANLATSHPHYDSAGN 180  
 QY 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 Db 181 ILNMGTSIVDKGRKRYKYLKIPSSVPEKKEKSCFPHLEVCSTPSRSILOPSYHSFGI 240  
 QY 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 Db 241 TENYIVLEQPFKLDIVKLAAYIRGVNASCLSPFKEDKTFHFYDRKTKKEVSTKRYT 300  
 QY 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 Db 301 DALVLYHHINAYEDGHNVPDIYAYRDNSTYDMFYLLKIDKDFEVNKKLSTPTCKRFVY 360  
 QY 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 Db 361 PLGYDKDAEVSGLVLRPTSATAVKEKGSITCOPELLEGIELPRVNYDYNGKKRYVY 420  
 QY 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 Db 421 ATEVQWSPVPKTKLAKLVQTKRYLHMGEHGMSEPTFVPSPAEEDSGVLTGVVSE 480  
 QY 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526  
 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526  
 Db 481 PNKAPFLILLDAKTFKELGRATVNEHMLDLHGMFIPONDLGAETE 526

RESULT 3  
 Q9ERN9 PRELIMINARY; PRT; 566 AA.

AC Q9ERN9  
 ID Q9ERN9  
 DT 01-MAR-2001 (TREMblrel. 16, created)  
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)  
 DE beta, beta-carotene 15,15'-dioxygenase (EC 1.13.11.21).  
 GN BC001 OR BC001  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RA Yan W., Jiang G.-F., Haeseleer F., Esuni N., Chang J., Kottmann M.,  
 Campochiaro M., Campochiaro P., Polgar W., K. Zook D.J.;  
 RT "Cloning and characterization of a human beta, beta carotene 15,15'-  
 dioxygenase that is highly expressed in the retinal pigment  
 epithelium."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBPRT databases.  
 DB EMBL: AF244899; AB15381.1;  
 DB MGP: MG11925923; Bcd01.  
 DB InterPro: IPR004294; RPF65.  
 DB Pfam: PF03055; RPF65; 1.  
 DB Dioxxygenase; Oxidoreductase;  
 SQ SEQUENCE 566 AA; 64852 MW; 2C4D9791B645670D 67664;



```

100 121 IDNCLINIMKGEFFATSTNTYIFKINPOTLEFTIKVYPRKVAANLASHPHTEACN 180
101 181 ILNMGTSIVKGRKTKVYIFKIPSSVPR-KEKKSCKPKHFWGTSIPSKSLGQSYHSHG 239
102 181 VLNMGTIVKGRKTKVYIFKIPATVEGKQKSPKHKTEFVCSIPSRSLTSSSYHSHG 240
103 240 ITENVYVEIFDQPEFLDILKAVIFGVNASTSLSHKREKTFHFHVEFFIKKVSIFKY 299
104 241 VTEENVYVEIFDQPEFLDILKAVIFGVNASTSLSHKREKTFHFHVEFFIKKVSIFKY 300
105 300 TVALVYHININAEFFSHVEDIVAVFVNSVIMFYKLKLKLEFVENNKLISIPKPV 359
106 301 TDAVAVFHHVNAVEEDGIVEFVIAVEDNSLYOIFVIANINQDFKNSKLSVPLPRFA 360
107 360 VPLDGYDADFVNSLVEL-PLSAFAVEKESSTVYQPEFLDILKAVIFGVNASTSLSHKREK 418
108 361 VPLHVNKNAEVEGNILKVAISTATALKKEEGGVYQPEFLDILKAVIFGVNASTSLSHKREK 420
109 419 VYATEVQWSPVPRKIAKLVNQTKEVLHMGEDHCPSEPIFVPSDAREDEGVVITGVV 478
110 421 VFATGVQWSPVPRKIAKLVNQTKEVLHMGEDHCPSEPIFVPSDAREDEGVVITGVV 480
111 479 SEPKAPFLDILDAKTEKELGRATVNVEMHDLHGPFPOND 520
112 481 TDPOKLPFLDILDAKSTELARASVDVDMHDLHGPFITDM 522

```

## RESULT 6

```

ID 090WH5 PRELIMINARY; PRT: 547 AA
AC 090WH5;
DT 01-OCT-2000 (TREMBL:rel. 15, Created)
DT 01-OCT-2000 (TREMBL:rel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBL:rel. 21, Last annotation update)
DE CDNA FLJ10730.115, clone NT2RP1001107
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN 11
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagaatsuma M., Hosoki T., Kato Y., Kodaira H., Sogawara M.,
RA Takahashi M., Gaba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto T., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,
RA Nishimura K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FBI-2000) to the EMBL/GenBank/DBJ databases
DK EMBL: AK001592; BAA91776.1;
DK InterPro: IPR004294; RPF65;
DK Pfam: PF04055; RPF65;
SO SEQUENCE 547 AA; 62579 MW; 394A18R01P6660EF C7664;

```

Query Match 74.68; Score 2107; DB 4; Length 547;

Best Local Similarity 70.38; Pred. No. 1,6e-148;

Matches 367; Conservative 96; Mismatches 57; Indels 2; Gaps 2;

```

QY 1 METIFNKNKEHPEPIKAEVQGOIPWLOGVLLRNPGMHTIGDKYNNHMFGLALHSF 60
DB 1 MDIIFGNKRKEQLEPRAKVATCKIPAMLOGTILRNPGMHTIGDESYNNHMFGLALHSF 60
QY 61 TFRNGEYVYSKYLRSQDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 120
DB 61 TFRNGEYVYSKYLRSQDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 120
QY 121 TDNCLINIMKGDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 180
DB 121 TDNCLINIMKGDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 180
QY 181 ILNMGTSIVKGRKTKVYIFKIPSSVPR-KEKKSCKPKHFWGTSIPSKSLGQSYHSHG 239
DB 181 ILNMGTSIVKGRKTKVYIFKIPSSVPR-KEKKSCKPKHFWGTSIPSKSLGQSYHSHG 239

```

```

DB 181 VLNMGTIVKGRKTKVYIFKIPATVEGKQKSPKHKTEFVCSIPSRSLTSSSYHSHG 240
QY 240 ITENVYVEIFDQPEFLDILKAVIFGVNASTSLSHKREKTFHFHVEFFIKKVSIFKY 299
DB 241 VTEENVYVEIFDQPEFLDILKAVIFGVNASTSLSHKREKTFHFHVEFFIKKVSIFKY 300
QY 300 TVALVYHININAEFFSHVEDIVAVFVNSVIMFYKLKLKLEFVENNKLISIPKPV 359
DB 301 TDAVAVFHHVNAVEEDGIVEFVIAVEDNSLYOIFVIANINQDFKNSKLSVPLPRFA 360
QY 360 VPLDGYDADFVNSLVEL-PLSAFAVEKESSTVYQPEFLDILKAVIFGVNASTSLSHKREK 418
DB 361 VPLHVNKNAEVEGNILKVAISTATALKKEEGGVYQPEFLDILKAVIFGVNASTSLSHKREK 420
QY 419 VYATEVQWSPVPRKIAKLVNQTKEVLHMGEDHCPSEPIFVPSDAREDEGVVITGVV 478
DB 421 VFATGVQWSPVPRKIAKLVNQTKEVLHMGEDHCPSEPIFVPSDAREDEGVVITGVV 480
QY 479 SEPKAPFLDILDAKTEKELGRATVNVEMHDLHGPFPOND 520
DB 481 TDPOKLPFLDILDAKSTELARASVDVDMHDLHGPFITDM 522

```

## RESULT 7

```

ID 090WH4 PRELIMINARY; PRT: 516 AA.
AC 090WH4;
DT 01-DEC-2001 (TREMBL:rel. 19, Created)
DT 01-DEC-2001 (TREMBL:rel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBL:rel. 20, Last annotation update)
DE Putative b,b'-carotene-15,15'-dioxygenase.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID:7955;
RN 11
RP SEQUENCE FROM N.A.
RA Kiefer C., Hessel S., Lampert J.M., Vogt K., Lederer M.O.,
RA Breithaupt D.E., vonLintig J.;
RT "Identification and characterization of a Mammalian enzyme catalyzing
RT the asymmetric oxidative cleavage of provitamin A."
PL J. Biol. Chem. 276:14119-14116(2001).
CP EMBL: AL199309; GAC37566.1;
DK InterPro: IPR004294; RPF65;
DK Pfam: PF04055; RPF65;
DK DIOXYGENASE.
SO SEQUENCE 516 AA; 58584 MW; DEDA47707B55C94F C7664;

```

Query Match 60.78; Score 1715; DB 13; Length 516;

Best Local Similarity 59.78; Pred. No. 2.2e-119;

Matches 310; Conservative 87; Mismatches 114; Indels 8; Gaps 7;

```

QY 1 METIFNKNKEHPEPIKAEVQGOIPWLOGVLLRNPGMHTIGDKYNNHMFGLALHSF 60
DB 1 MDIIFGNKRKEQLEPRAKVATCKIPAMLOGTILRNPGMHTIGDESYNNHMFGLALHSF 60
QY 61 TFRNGEYVYSKYLRSQDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 120
DB 61 TFRNGEYVYSKYLRSQDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 120
QY 121 TDNCLINIMKGDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 180
DB 121 TDNCLINIMKGDYVATSETNIRKIDPOTLETLQKVDYKVVANVATSHPHDSAGN 180
QY 181 ILNMGTSIVKGRKTKVYIFKIPSSVPR-KEKKSCKPKHFWGTSIPSKSLGQSYHSHG 239
DB 181 ILNMGTSIVKGRKTKVYIFKIPSSVPR-KEKKSCKPKHFWGTSIPSKSLGQSYHSHG 239

```

[illegible]

```

Query Match: 39.5%, Score 115.07, DB 4; Length 545;
Best Local Similarity: 44.54, Pct Id: 9.26-75;
Matches: 233, Conservative: 96; Mismatches: 192; Indels: 23; Gaps: 9

QY 10 EEHPEPIKAEVQGLDPTWLDGVLKRGPMHIGDTKYNHMEDGIALHSHTEFNKGEVY 69
TH 28 EEAPEHLSAAYWHFEKRWLNGSLTIGKKEPEKTKYHMLDWMALHTEFNAGTIVY 87
QY 70 PSATYPSPTATNENIENFVVSSEFTMAITPTSTYHFKAKS YLSTTEFTPTCLIN 127
TH 88 KSTHEDSLDGIYKANSANRRIVLSHGLALDELALANVFEFEKSKTELEFQNAAMRLDINVN 147
QY 128 IMKQGLVAVATSEFNELEKIPDQETPTTEKGVNRYAVANLATSHPHDSAGNTIIMMIS 187
TH 148 YVYKRYKLYLCEITNRMNKKVLEIELTEKRYWSELNANALAMHHEHGLAVNRNS 207
QY 188 IVKQPTVYVLEKIPSSVPEKEKKKSCMHLEVWSLSPSLDQYVHSGTENTVIF 247
TH 208 FCHYVC-PSYKVFIPF---PEKYNQPLDINGVAVTISLSTKQKSLSYHSGKTPRYLIF 263
QY 248 LEDEPKFLDVKLATAVLEGVNMMASLELSEKELKTEHGVADREKREVSATRYTALVIH 307
TH 264 LEGLFQKRLKIAISKIPKAFKSLSTWFLQNIHFWHAFKFLQGLDLYVSHVFTYQ 433
QY 308 HINAYFEGHVPDIIAAYPDNSLYMPYIKKLLK ----GREVNKKTSITPCKEYVPL 362
TH 424 EIHAFHFGGVVITGCTEPEKFTLVYTGQRLFRAGDELLFYHNAKAGCTLRREVLPL 381
QY 363 QYTFLEAVSNLKLPLTSALVKEKQSTYVQPELGE ----GTEFKNVHLYVNS 413
TH 382 NVSLINAFEGDNLSPISYLSASAVKAKANDIWCSEHNLHQDELEKQGIPEQPIAYVDFNS 441

```

QY 414 KKRYVATEVQWSPVPTKIAKINVOETREVLHMGCDHCEWSPSPITVPSDAREPBGVYL 473  
 DB 442 KKHHFFGGGGR-HLVGDSLLKVDVYNKILKVKREGLTFPSEPVVAPAGTNEEDGVYL 500  
 QY 474 TCVVVSPNKAPELLLLDAKTKRELGATVNMHMLDGMETIP 517  
 DB 501 SVVITPNQNESNPLVLDAKNPEELGRAEVPMYGPFGHTIP 544

## RESULT 10

Q99NFI PRELIMINARY: PRT: 532 AA.  
 ID Q99NFI  
 AC Q99NFI  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE B.D-carotene 9,10'-dioxygenase.  
 GN BCD02 OR B-DIOX-11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE=21216714; PubMed=11278918;  
 RA Kleier G., Hessel S., Lampert J.M., Vogt K., Lederer M.O.,  
 RA Breithaupt D.E., von Linling J.;  
 RT "Identification and characterization of a mammalian enzyme catalyzing  
 RT the asymmetric oxidative cleavage of provitamin A."  
 RL J. Biol. Chem. 276:14110-14116(2001).  
 DR EMBL: AJ230492; F0628626.1;  
 DR M00: M01:2177469; Bcd02.  
 DR InterPro: IPR004294, RPE65.  
 DR Pfam: PF03055; RPE65; 1.  
 KW Dioxygenase.  
 SO SEQUENCE 532 AA; 60142 MW; 7461AD5A53F86A3 CRC64;

Query Match 39.4%; Score 1112.5; DB 11; Length 532;  
 Best Local Similarity 43.5%; Pred No. 1.5; 74;  
 Matches 228; Conservative 94; Mismatches 175; Indels 27; Gaps 10;

QY 10 EEHPEPIKAEVQGLPTWLOGVLLHNGPDMHTIGTKYHNMFDGLALLHSTFFKNGVY 69  
 DB 19 EEFLSAVASARVRGHPEPMNGVLLRVGPKPEFGKURVHMDGALLHQEFMEGTVY 78  
 QY 70 RSKYLSDDYVNCNIEANRIVSFHGTMAVDPCKNIFAKAFSYLSHTIPFEDNCLIM 129  
 DB 79 KSKFLQSDTYKANSAGRIIVISFQTLALPDCKSTFERKMRFP--PTMTDNTNVFV 136  
 QY 130 KTDIDVYATSENFIRKIDPQLETLKVDYSKYVAVNLATISHPHYSAGNLLNGTSIV 189  
 DB 137 QVGGDYVMSTENFMKRVDIEMLETERKVDMSKFLAVNGATAHPRYDGTAYNMGNSYG 196  
 QY 190 DKCRITVYLFKLISSVPEKSKCKHLEVCSSIPSRSLDPSYHSFGTENTVIVLE 249  
 DB 197 PRG-SYNNITFPV---PKRKEPETIHGAGVLCSTASTKMKRSTYSHSGMKNTIIVE 252  
 QY 250 QPEKIDIVKLATAYIGVNWASCLSFHKEDKTFPHFVDRKTKREYSTFYDALVYHHI 309  
 DB 253 QIPKMKIMKILTSKIKGKPFADGISWEPOYNTFRFVVDKHTQLLRPGMYSPBLTYHQI 312  
 QY 310 NAVEEHGHVFDLVATKQNSLYDMFLK-----LKDREVNKKLSTIPCKRFVPL 362  
 DB 313 NAVEHDCGIVLIDCGDGRSLDIYOLQNIKRAKGFIDQVYPL--KANSFP--PPVPL 368  
 QY 363 QYDKDAVGSNLVKLP-TSATAVKEKDSIFQPELCE-----GIELPRVND-YNG 413  
 DB 369 IAVVDAAIKKNLSPLSYSSASAVAGGTGTWCSPTNIHHPFI FREGGIEFPQIYNGPNS 428  
 QY 414 KKRYVATEVQWSPVPTKIAKINVOETREVLHMGCDHCEWSPSPITVPSDAREPBGVYL 473

DB 429 KKSHFVGGGGR-HLVGDSLLKVDVYNKILKVKREGLTFPSEPVVAPAGTNEEDGVYL 487  
 QY 474 TCVVVSPNKAPELLLLDAKTKRELGATVNMHMLDGMETIP 517  
 DB 488 SVVITPNQNESNPLVLDAKNPEELGRAEVPMYGPFGHTIP 544

## RESULT 11

Q96JY5 PRELIMINARY: PRT: 539 AA.  
 ID Q96JY5  
 AC Q96JY5  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE CONA FLJ14895 fis. clone JACR100449, weakly similar to B.taurus  
 DE retinal pigment epithelial membrane receptor p63.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Isogai T., Gao F., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto T., Wakamatsu A.,  
 RA Nakamura Y., Nishihara K., Masuda Y., Sasaki N.;  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK827801; HA855479.1;  
 DR InterPro: IPR004294; RPE65.  
 DR Pfam: PF03055; RPE65; 1.  
 SW SEQUENCE 539 AA; 60826 MW; 64346112031B861 CRC64;

Query Match 39.0%; Score 1102.5; DB 4; Length 539;  
 Best Local Similarity 44.3%; Pred No. 8.4; 74;  
 Matches 232; Conservative 85; Mismatches 179; Indels 29; Gaps 10;

QY 10 EEHPEPIKAEVQGLPTWLOGVLLHNGPDMHTIGTKYHNMFDGLALLHSTFFKNGVY 69  
 DB 28 EEMPEGASVVMVSHFPEWNLSELRDPEKEFEKLVYNNHMDGALLHQEFMKAGVYL 87  
 QY 70 RSKYLSDDYVNCNIEANRIVSFHGTMAVDPCKNIFAKAFSYLSHTIPFEDNCLIM 127  
 DB 88 RSKFLQSDTYKANSAGRIIVISFQTLALPDCKSTFERKMRFP--PTMTDNTNVFV 147  
 QY 128 IMPTDGYVATSENFIRKIDPQLETLKVDYSKYVAVNLATISHPHYSAGNLLNGTS 187  
 DB 148 YVYKCDYVLCETINRMNVDFEELTFTEKRWMSFVAVNATGABHIDQGTAVYNGNS 207  
 QY 188 IVDKRTVYVLEFIPSSVPEKSKCKHLEVCSSIPSRSLDPSYHSFGTENTVIVLE 247  
 DB 208 ECLYGG-ESKRVLPV---PEKVDLDELHGVVGLSTASIKRKGSTYSHSGMKNTIIVE 264  
 QY 248 IPEPKIDIVKLATAYIGVNWASCLSFHKEDKTFPHFVDRKTKREYSTFYDALVYHHI 307  
 DB 264 IEPPLKMKIMKILTSKIKGKPFADGISWEPOYNTFRFVVDKHTQLLRPGMYSPBLTYHQI 324  
 QY 308 HINAEDEGHVFDLVATKQNSLYDMFLK-----LKDREVNKKLSTIPCKRFVPL 362  
 DB 324 QINAPFLQGV-VIIDDGQDNGSPLTFYVYQNTIFKASGIDLDVYHNSAVKSP--KKVPL 381  
 QY 363 QYDKDAVGSNLVKLP-TSATAVKEKDSIFQPELCE-----GIELPRVND-YNG 413  
 DB 382 NVSLNAPEDDNLSPLSYTSASAVKADDTIWESHNLNGEDELKRGKGFEPQVYDPSG 441  
 QY 414 KKRYVATEVQWSPVPTKIAKINVOETREVLHMGCDHCEWSPSPITVPSDAREPBGVYL 473  
 DB 442 KKHHFFGGGGR-HLVGDSLLKVDVYNKILKVKREGLTFPSEPVVAPAGTNEEDGVYL 494  
 QY 474 TCVVVSPNKAPELLLLDAKTKRELGATVNMHMLDGMETIP 517  
 DB 495 SVVITPNQNESNPLVLDAKNPEELGRAEVPMYGPFGHTIP 548



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;  
 OX NCBI\_TaxID=9606;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RETINAL PIGMENT EPITHELIUM;  
 RX MEDLINE=9535966, PubMed=763413;  
 RA Nicoletti A., Wong D.J., Kawase K., Gibson L.H., Yang-Peng T.L.,  
 RA Richards J.E., Thompson D.A.;  
 RT "Molecular characterization of the human gene encoding an abundant 61  
 RT kDa protein specific to the retinal pigment epithelium";  
 RL Hum. Mol. Genet. 4:641-649(1995).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RA Morimura H.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U18991; AAA99012.1; -;  
 DR EMBL: U20510; AAC14586.1; -;  
 DR EMBL: U20476; AAC14586.1; JOINED.  
 DR EMBL: U20477; AAC14586.1; JOINED.  
 DR EMBL: U20478; AAC14586.1; JOINED.  
 DR EMBL: U20479; AAC14586.1; JOINED.  
 DR EMBL: U20481; AAC14586.1; JOINED.  
 DR EMBL: U20482; AAC14586.1; JOINED.  
 DR EMBL: U20484; AAC14586.1; JOINED.  
 DR EMBL: U20485; AAC14586.1; JOINED.  
 DR EMBL: U20486; AAC14586.1; JOINED.  
 DR EMBL: AF039868; AAC19660.1; -;  
 DR EMBL: AF039865; AAC19660.1; JOINED.  
 DR EMBL: AF039856; AAC19660.1; JOINED.  
 DR EMBL: AF039857; AAC19660.1; JOINED.  
 DR EMBL: AF039858; AAC19660.1; JOINED.  
 DR EMBL: AF039859; AAC19660.1; JOINED.  
 DR EMBL: AF039860; AAC19660.1; JOINED.  
 DR EMBL: AF039861; AAC19660.1; JOINED.  
 DR EMBL: AF039862; AAC19660.1; JOINED.  
 DR EMBL: AF039863; AAC19660.1; JOINED.  
 DR EMBL: AF039864; AAC19660.1; JOINED.  
 DR EMBL: AF039865; AAC19660.1; JOINED.  
 DR EMBL: AF039866; AAC19660.1; JOINED.  
 DR EMBL: AF039867; AAC19660.1; JOINED.  
 DR InterPro: IPR004294; PPF65;  
 DR Pfam: PPF03055; PPF65; 1  
 NC SEQUENCE 533 AA; 60947 MW; 7193C93F325798D CRC64;  
 Query Match 38.1%; Score 1075; DB 4; Length 533;  
 Best local similarity 42.2%; Pred. No. 9, 26-72;  
 Matches 222; Conservative 93; Mismatches 187; Indels 24; Gaps 12;  
 QY 2 EITFNKKEHPPIAEVGGOLPTWLGVLNRPQMTTGDTRYNMFGDLAHSFT 61  
 DB 17 ETV-----ELSTPLTAHTGRIPLMLTSLKCGLEFVSEPFYHFLDQALLHKFD 71  
 QY 62 FRKNEYVRSKYRSTQYNINLEAFIVSESGIMAYRQCKNIJAKAKSYSHIIPET 121  
 DB 72 FRKHVHYHPPPTPTDAVYVAMTEKPIVITTEGTCAFPKCKNITSPPFSYRGV--EVT 129  
 QY 122 ENCLINIMKTGDVYATSETNFKIDPQTELETLDKVDYSKYVAANLATSHPHDSAGNI 181  
 DB 130 DNALVINYVPGFYVACTENFTKINPTELETIKQVLDICNVSVNGATAHPIHINDGIV 189  
 QY 182 LNMGTSLVKGRTKRYLFRIPSSVPEKSKCKSPKLEVVCGISPSKSLDSYHSFOIT 241  
 DB 190 YNIGDCGKNFSIAYNIKIP--PLQADKEPDISKSEIIVQFPCSDPKPSYVHSFGLT 246  
 QY 242 ENYIVFLEQPKLDIYKLTAV--IRGVNASCISPEHKEKTPWFHVDKRTKKEVSTRKYT 300  
 DB 247 PNIVIVETIVKINLEKFLSSWSIMGANMDFESNEMTGVALHIAKKRKKYLLNKKRT 306  
 QY 401 DALVIVHINAVFEDGHVVDIVAYRD--NSLYDMFLYKLLDKDFE---VNNKLITS1PTCK 356  
 DB 307 SPNPLPHHINITYEDNGFLIVDLICQMKGFEEVYNYIYIAMLRENWEVAKNAKAPQPEVR 366

QY 357 PFVVLGYCKDAEVSNIKLP--TSAFAKEMKSTYQPIELCEG---LELHVNYD 410  
 DB 367 RYVPLINDK-ADTGKNLITLPPITVATLLSNETTWLEPEVLSAPRQVAFYQINVK 425  
 QY 411 YNCKRYKRYVAIVGVSWPEPIALINVS--GVHMDSHLWSEPTVPSAKKELGG 470  
 DB 426 YCKRYVAYVGDLSI NHH-VPHQCKI NVAKIKELWWQFHSSEEDFVSDIALERDDG 484  
 QY 471 VVLGVVSEPNKAP--FLILLAKTKELCKRATVANNEMIDILKMP 515  
 DB 485 VLVSVVSPGAGQPAVILLIINKRDISVAPAPAVETINIVPHOLF 540  
 RESULT 15  
 ID 070276 PRELIMINARY; PRT: 533 AA.  
 AC 070276;  
 DT 01-AUG-1998 (TREMURel. 07, Created)  
 DT 01-AUG-1998 (TREMURel. 07, last sequence update)  
 DT 01-DEC-2001 (TREMURel. 19, last annotation update)  
 DE Retinal pigment epithelium specific protein.  
 GN RPE65.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;  
 OX NCBI\_TaxID=10116;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MISTAR; TISSUE=RETINA;  
 RX MEDLINE=98171299; PubMed=9512345;  
 RA Mares G., LeDucq P., Kucharczak J., Pages A., Schmitt Bernard C.P.,  
 RA Hamel C.P.;  
 RT "Rat messenger RNA for the retinal pigment epithelium specific protein  
 RT RPE65 gradually accumulates in two weeks from late embryonic days.";  
 RL FEBS Lett. 423:133-137(1998).  
 DR EMBL: AF035673; AAC40059.1; -;  
 DR InterPro: IPR004294; PPF65;  
 DR Pfam: PPF03055; PPF65; 1  
 NC SEQUENCE 533 AA; 60989 MW; EB0712646CF52693 CRC64;  
 Query Match 37.9%; Score 1072; DB 11; Length 533;  
 Best local similarity 42.6%; Pred. No. 1, 60-71;  
 Matches 224; Conservative 89; Mismatches 193; Indels 24; Gaps 12;  
 QY 2 EITFNKKEHPPIAEVGGOLPTWLGVLNRPQMTTGDTRYNMFGDLAHSFT 61  
 DB 17 ETV-----ELSTPLTAHTGRIPLMLTSLKCGLEFVSEPFYHFLDQALLHKFD 71  
 QY 62 FRKNEYVRSKYRSTQYNINLEAFIVSESGIMAYRQCKNIJAKAKSYSHIIPET 121  
 DB 72 FRKHVHYHPPPTPTDAVYVAMTEKPIVITTEGTCAFPKCKNITSPPFSYRGV--EVT 129  
 QY 122 ENCLINIMKTGDVYATSETNFKIDPQTELETLDKVDYSKYVAANLATSHPHDSAGNI 181  
 DB 130 DNALVINYVPGFYVACTENFTKINPTELETIKQVLDICNVSVNGATAHPIHINDGIV 189  
 QY 182 LNMGTSLVKGRTKRYLFRIPSSVPEKSKCKSPKLEVVCGISPSKSLDSYHSFOIT 241  
 DB 190 YNIGDCGKNFSIAYNIKIP--PLQADKEPDISKSEIIVQFPCSDPKPSYVHSFGLT 246  
 QY 242 ENYIVFLEQPKLDIYKLTAV--IRGVNASCISPEHKEKTPWFHVDKRTKKEVSTRKYT 300  
 DB 247 PNIVIVETIVKINLEKFLSSWSIMGANMDFESNEMTGVALHIAKKRKKYLLNKKRT 306  
 QY 357 PFVVLGYVNDVAVSNIVYIP--TCAVAVFPFVSQVYQPELPHG---LELHVNYD 410  
 DB 367 RYVPLPHHINITYEDNGFLIVDLICQMKGFEEVYNYIYIAMLRENWEVAKNAKAPQPEVR 425







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Inc

OM nucleotide sequence search, using SW model

Run on: July 15, 2003, 16:11:50 : Search time 0.15 seconds  
(without alignments)

10895 749 Million cell updates/sec

Title: US-10-053-192-2

Perfect score: 3111  
Sequence: 1 cggatccactagtaacgggc ..... aaaaaaaaaaaaaaaaaa 3111

Scoring table: IDENTITY\_MTC  
Gapop 10 0, Gapext 1 0

Searched: 2160339 seqs, 115399159 positions

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N\_Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	3111	100.0	3111	21	AAAS3888
2	892.4	25.8	2437	22	AAH14474
3	577.2	18.6	1773	22	AAH14170
4	333.4	10.7	809	22	AAH06478
5	278.8	9.0	2051	22	AAF58536
6	272.4	8.8	1934	22	AAH14870
7	264.8	8.5	1724	22	AAAS07196
8	261.4	8.4	1855	22	AAF79560
9					AAAS07194

10	250.8	8.1	2134	22	AAAS7195	DNA encoding beta
11	245.4	7.9	1599	14	AAAS4070	Human retinol bind
12	145.6	4.4	825	17	AAH76639	Human cDNA clone (
13	104.6	3.4	1265	14	AAAS3074	Human retinol bind
14	84.4	2.7	1978	23	ABL29861	Drosophila melanog
15	84.4	2.7	4837	22	AAAS7180	DNA encoding Dros
16	84.4	2.7	2037	22	AAH42156	Human cDNA clone
17	67.2	2.2	4414	23	ABL27850	Drosophila melanog
18	60.2	1.9	927	21	AAAS7854	DNA encoding novel
19	60.2	1.9	432	21	AAAS7172	Human secreted pro
20	57.6	1.9	5000	24	ABL55643	AMEV genome fragm
21	50.8	1.6	15548	24	ABL34155	Human genome fragm
22	48.2	1.5	1713	22	AAH19635	Arabidopsis thalia
23	47.6	1.5	73334	22	ABL92319	Chemically treated
24	47.6	1.5	73334	24	ABL34125	Human immune syste
25	47.2	1.5	15479	24	ABK39965	Human cDNA clone (
26	47.2	1.5	6175	24	ABL33307	Human immune syste
27	46.6	1.5	6071	24	ABL33325	Human immune syste
28	46.6	1.5	6071	24	AAAS1076	Human gene regulat
29	46.6	1.5	6650	24	ABL32548	Human immune syste
30	46.4	1.5	3037	24	ABK33094	Human RNA for stag
31	46.4	1.5	9219	22	AAAS4607	Tumour suppressor
32	46.4	1.5	9219	24	ABL34426	Human immune syste
33	46.2	1.5	591	22	AAH12584	Human cDNA clone (
34	46.2	1.5	6337	22	AAAS4650	Tumour suppressor
35	46.2	1.5	6337	22	ABL33311	Human immune syste
36	45.8	1.5	1791	22	AAH19634	Arabidopsis thalia
37	45.8	1.5	16724	24	ABL70260	Chemically treated
38	45.8	1.5	16724	24	ABL33091	Human immune syste
39	45.8	1.5	16724	24	ABL34577	Human immune syste
40	45.6	1.5	5123	24	ABL33921	Human immune syste
41	45.2	1.5	13038	24	ABL33074	Human immune syste
42	45	1.4	5523	24	ABL33155	Human immune syste
43	45	1.4	5829	22	AAAS4683	Chemically treated
44	45	1.4	5829	22	AAAS46425	Tumour suppressor
45	45	1.4	5829	24	ABK28325	DNA transcription

#### ALIGNMENTS

RESULT 1	AAAS3888	standard; DNA: 3111 BP.
ID	AAAS3888:	
XX	03-JAN-2001	(first entry)
AC	AAAS3888:	
XX		
DT	03-JAN-2001	(first entry)
XX		
DE	Beta, beta-carotene-15,15-dioxygenase coding sequence.	
XX		
KW	Beta, beta-carotene-15,15-dioxygenase; vitamin A; beta-carotene;	
KW	lutein; xanthophyll; phytyl; phytyl; phytyl; phytyl; phytyl;	
KW	ophthalmological disorder; antibody; detection; quantification;	
KW	treatment; therapy; ds.	
XX		
OS	Gallus gallus.	
XX		
PH	KEY	Location/Qualifiers
FT	CDS	133..1713
FT		/tag a
FT		/product Beta, beta-carotene-15,15-dioxygenase
XX		
PN	EP1031627-A1.	
XX		
PT	30-AUG-2000.	
XX		
XX	17-FEB-2000.	200001010289.
XX		
XX	22-FEB-1999.	99010103482.
XX		
PA	(HOF) HOFMAN, LA FOUTH & CO AG F.	
XX		













PA (INCYTE GENOMICS INC.  
 XX Au-Young J, Bandman O, Tang YT, Yue H, Azimzai Y, Burford N;  
 PI Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;  
 XX WPI; 2001-168554/17.  
 DR P-PSDB; AAB68872.  
 XX  
 PT Novel receptors and associated proteins for diagnosis and treatment of  
 PT neurological disorders, immunological disorders including autoimmune/  
 PT inflammatory disorders and cell proliferative disorders such as cancer  
 XX  
 PS Claim 5; Page 115-116; 128pp. English.  
 XX  
 CC The present sequence encodes a human RETAP (receptors and associated  
 CC proteins) polypeptide. RETAP polynucleotides and polypeptides are useful  
 CC in the diagnosis, treatment and prevention of neurological disorders  
 CC such as stroke, Alzheimer's disease, Pick's disease, Huntington's  
 CC disease, dementia, Parkinson's disease, Down's syndrome, amyotrophic  
 CC lateral sclerosis, multiple sclerosis, bacterial and viral meningitis,  
 CC CBD (Creutzfeldt-Jakob disease), GSS (Gerstmann-Sträussler-Scheinker  
 CC syndrome), immunological disorders, including autoimmune/inflammatory  
 CC disorders such as AIDS, Digeorge's syndrome, severe combined  
 CC immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's  
 CC disease, Addison's disease, autoimmune thyroiditis, Crohn's disease,  
 CC diabetes mellitus, good pasture's syndrome, gout, Grave's diseases,  
 CC Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral,  
 CC bacterial, fungal, parasitic, protozoal, and helminthic infections; and  
 CC cell proliferation disorders such as arteriosclerosis, atherosclerosis,  
 CC cirrhosis, hepatitis and cancer.  
 XX  
 SO Sequence 2051 BP; 578 A; 412 C; 508 G; 553 T; 0 other;  
 Query Match 9.0%; Score 278.8; DB 22; Length 2051;  
 Best Local Similarity 51.5%; Pred. No. 2.8e-58;  
 Matches 804; Conservative 0; Mismatches 707; Indels 51; Gaps 5;  
 QY 160 GAAGAGCATCCAGAGGCTATTAAGTCTGAGTGTAAAGTCACTGATGCTGTTGTTAA 419  
 DB 278 GAAGAGGCTCCAGAGGCTATTAAGTCTGAGTGTAAAGTCACTGATGCTGTTGTTAA 337  
 QY 220 GGGGATATTTTGAATGAGGCTAGGAGTAAATGAGTAAATGAGTAAATGAGTAAATG 279  
 DB 338 GGGGATATTTTGAATGAGGCTAGGAGTAAATGAGTAAATGAGTAAATGAGTAAATG 397  
 QY 280 TGGTGTAGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 339  
 DB 398 TGGTGTAGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 457  
 QY 340 AGAAGTAGTACCTGCGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACG 399  
 DB 458 AGAAGTAGTACCTGCGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACG 517  
 QY 400 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453  
 DB 518 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 577  
 QY 454 AAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 513  
 DB 578 AAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 637  
 QY 514 ATTATGAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 573  
 DB 638 ATTATGAAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 697  
 QY 574 GATCCAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 633  
 DB 698 GATCCAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 757  
 QY 634 TTGCTACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 693  
 DB 758 TTGCTACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 817

QY 694 ATTGTGATTAAG 753  
 DB 818 TTGCTACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 865  
 QY 754 AAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 813  
 DB 866 AAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 925  
 QY 814 TGGTGTAGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 873  
 DB 926 TGGTGTAGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 985  
 QY 874 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933  
 DB 986 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1045  
 QY 934 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 993  
 DB 1046 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1105  
 QY 994 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1053  
 DB 1106 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1165  
 QY 1054 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1113  
 DB 1166 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1225  
 QY 1114 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1164  
 DB 1226 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1285  
 QY 1165 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1224  
 DB 1286 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1345  
 QY 1225 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1281  
 DB 1346 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1405  
 QY 1282 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1349  
 DB 1406 ATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1465  
 QY 1330 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1380  
 DB 1466 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1525  
 QY 1381 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1440  
 DB 1526 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1585  
 QY 1441 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1500  
 DB 1586 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1645  
 QY 1501 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1560  
 DB 1646 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1705  
 QY 1621 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1680  
 DB 1766 AAGTGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1825  
 QY 1681 CC 1682  
 DB 1826 CC 1827

RESULT 6  
 AAH14870  
 ID AAH14870 standard; cDNA: 1983 bp.  
 AC AAH14870;  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12717.  
 XX  
 KM Human: primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 PN EP1074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI) HELIX RES INST.  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 TS Claim 8; SEQ ID 12717; 2537pp + CD ROM, English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one or  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH3628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to  
 CC AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1983 BP; 583 A; 384 C; 493 G; 523 T; 0 other;  
 Query Match 8.88; Score 272.8; DB 22; Length 1983;  
 Best Local Similarity 51.18; Pred. No. 8; 2e-57;  
 Matches 789; Conservative 0; Mismatches 722; Indels 33; Gaps 5;  
 QY 160 GAAATGATTCACAGCCGATATAAAATGTGAGTGGAGAGTTCAGTTCGCACTTGTGCA 219  
 DB 184 GAAATGATTCACAGCCGATATAAAATGTGAGTGGAGAGTTCAGTTCGCACTTGTGCA 243

QY 229 GGGGACCTTCTGCGAAATGGCGGACAGTGTATGAAATAGGCGATATATATATATATAT 279  
 DB 244 GATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 404  
 QY 280 TGGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 439  
 DB 304 TGGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 464  
 QY 340 ACAAATGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 499  
 DB 364 ACAAATGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 424  
 QY 400 GTCGTGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 454  
 DB 424 ATCTTGAATATTTGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 484  
 QY 454 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 514  
 DB 484 ATCTTGAATATTTGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 544  
 QY 514 ATTTGAAATGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 574  
 DB 544 TATGTCGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 604  
 QY 574 GATTCGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 634  
 DB 604 GATTCGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 664  
 QY 634 TTGCGACCTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 694  
 DB 664 GATTCGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 724  
 QY 694 ATTTGATTAAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 754  
 DB 724 TTTT-----GCGATATGCTTCGCTTATGACCTTATTTGCTTATTTGCTTATTTGCTT 774  
 QY 754 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 814  
 DB 774 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 834  
 QY 814 TCCCTGCTTCAAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 874  
 DB 834 GAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 894  
 QY 874 ATGAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 934  
 DB 894 ATGAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 954  
 QY 934 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 994  
 DB 954 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1014  
 QY 994 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1054  
 DB 1014 AAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1074  
 QY 1054 CACATTAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1104  
 DB 1074 CACATTAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1144  
 QY 1104 TACGAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1164  
 DB 1132 GGAAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1194  
 QY 1165 GGAAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1224  
 DB 1192 CAGGTCATTAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1254  
 QY 1225 GAGGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1284  
 DB 1252 AGTTGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1314



634 TTGGACCTTCACACACACATGACAGTGGTGGAAATATCTCAGATGGGACTTCA 693  
 695 GAAATACCTGACATTTATTATATATATATATATATATATATATATATATAT 654  
 696 ATTTGATTAAGGACACAAATATGCTCTTAAAGATGCTCTGCTGACAGAA 753  
 697 TTTT-----GAGCAATAGGCTTCTGATTAAGGCTATGCTGCTGCTGACAG 702  
 754 AAGGAAAGAAATATCTTTTAAACACTGAACTAGATGCTGCTGCTGCTGCTG 813  
 703 GAGGAGACCTTGGGAGATATATATATATATATATATATATATATATATAT 762  
 814 TCCCTGCTCAACAGCTTACACAGCTTGGATGACAAATATATATGCTGCTG 873  
 763 GAGAAAGGAAACCTTTCTTACTACTATAGCTTTGAAATGAAAGAACTATAT 822  
 874 ATAGAGCAGCCATTTAACTGATATGCTCAACTGCAACTGCTGCTGCTGCTG 933  
 823 ATTTAACTAACTTAAATATATATATATATATATATATATATATATATAT 882  
 934 AACTGAGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 993  
 883 GCTTTTACATGATATATATATATATATATATATATATATATATATATAT 942  
 994 AAGAAAGGAAAGAAAGATATATATATATATATATATATATATATATATAT 1053  
 943 AAGGACCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1002  
 1054 CACATTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1104  
 1003 CAATTAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1062  
 1105 TACAGAGCAATATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1164  
 1063 GGAAGACCCCTAGAGTTTACCACTTACAGATCTGAGAGGCTGGAGGCTGAT 1122  
 1165 GTGAAGCAGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1224  
 1123 GAGTCTCAATTAATGAGAGCAATCTTCTTCTTCTTCTTCTTCTTCTTCT 1182  
 1225 GACAGAGTCTGAGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1281  
 1183 AGTTGATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1242  
 1282 GTAAAGGAAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1329  
 1243 GTGAAGCAGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1302  
 1330 -----GAGGATAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 1303 GAAAGGAAAGGAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1442  
 1381 AAGTATGCTATGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAG 1440  
 1363 TATCATTTCTTTTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1422  
 1441 AATTTCTAAGGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
 1423 GATCTGCTGATTAAGACACTGAGAGTCTTCTTCTTCTTCTTCTTCTTCT 1482  
 1501 AATTTCTGCTGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAG 1560  
 1483 GTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1542  
 1561 GTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620  
 1543 ATCATCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 1602  
 1621 CATTTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 1680  
 1603 GATTTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 1662

QY 1681 CC 1682  
 DB 1663 CC 1664  
 RESULT 8  
 AAF79560  
 ID AAF79560 standard; CUNA: 1724 BP.  
 XX  
 AC AAF79560;  
 DE 29-MAY-2001 (first entry)  
 XX  
 DE Canine RPE65 CDNA.  
 XX  
 KW Canine, dog, RPE65, canine retinal pigment epithelium 65; RPE65;  
 XX congenital stationary night blindness; CSNB; mutation detection; ss.  
 XX  
 OS Canis familiaris.  
 XX  
 PN US6201114-B1.  
 PD 13-MAR-2001.  
 PF 30-AUG-1999; 9905-0385259.  
 PR 06-OCT-1998; 98US-0103219.  
 PA (CORR.) CORNELL RES FOUND INC.  
 XX  
 PI Aguirre GD, Acland GM, Ray K;  
 DR WPI: 2001-265168/27.  
 DR P-PSDB: AAB73955.  
 PT Novel nucleic acid molecule encoding canine retinal pigment epithelium  
 PT 65, where presence of mutation in one or both alleles is indicative of  
 PT blindness.  
 PT  
 PS  
 PS Claim 1; Fig 1A-1B; 25pp; English.  
 CC The present sequence encodes canine retinal pigment epithelium 65 (RPE65)  
 CC which has a sequence of 519 amino acids defined in the specification.  
 CC The gene is useful for identifying dogs which are genetically  
 CC normal, or are carriers of, or affected with congenital stationary  
 CC night blindness (CSNB). Carriers for CSNB are identified through the  
 CC detection of an AAG deletion at nucleotides 487-490 in the  
 CC canine RPE65 gene. This allows a breeder to eliminate the carrier from  
 CC the breeding stock or to breed carriers with genetically normal dogs.  
 CC  
 SQ Sequence 1724 BP; 465 A; 421 G; 294 C; 444 T; 0 other;  
 Query Match 8.5%; Score 264.8; DB 22; Length 1724;  
 Best Local Similarity 52.1%; Prod. No. 7.2e-55;  
 Matches 81; Conservative 792; Mismatches 45; Gaps 8;  
 QY 153 AAGGAAAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 212  
 DB 78 AAGGAAAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 137  
 QY 213 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 272  
 DB 138 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 197  
 QY 273 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 342  
 DB 198 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 257  
 QY 343 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 392  
 DB 258 GATGAGAGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 317





XX	oli seed/maize, chick pea, diagnostic, therapeutic, lilyflower, zebrafish,
KM	retinoid/vitamin A deficiency; beta-diox II; transgenic nutrition;
KW	carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop; ss.
OS	Brachydanio rerio.
XX	
XX	Key Location/Qualifiers
FT	CDS 29..1678
FT	/tag= "a"
FT	/product= "zebra-2"
XX	/note= "Beta-carotene dioxygenase (beta-diox II)"
XX	W0200148163-A2.
PD	05-JUL-2001.
PF	27-DEC-2000, 2000MO-EPI3273.
PR	24-DEC-1999; 99EP-0125895.
FR	20 MAR-2000, 2000EF 0105822.
PA	(GSEF-) GENOVATION PLANT/ENRICHMENT/5TE GMPH
PI	Von Lintig J, Vogt K:
XX	WPI: 2001-425657/45
DR	P-PDB: AAU04293.
XX	
PT	Novel isolated mouse, human, zebrafish beta-carotene dioxygenase
FT	(beta-diox II) protein that cleaves beta carotene and lycopene to yield
XX	beta-apo-carotenal and beta ionone, and apolycenals, respectively -
PS	Claim 8; Page 106-109; 116pp: English.
XX	
CC	The sequence represents the coding sequence of zebra-2, beta-carotene
CC	dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene
CC	and lycopene to form beta-apocaroetenal and beta-ionone, and
CC	apolycenals, respectively. The DNA is useful for transforming grain
CC	seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;
CC	edible seeds or seeds with edible parts e.g. chick-peas; potatoes;
CC	carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful
CC	for diagnostic and/or therapeutic purposes and for preparing antisense or
CC	ribozyme type therapeutic agents and for detecting any abnormality of
CC	endogenous beta-diox II. The beta-diox II specific antisense
CC	oligonucleotides derived from the DNA sequence are useful for dose
CC	response studies in relevant models of retinoid/vitamin A deficiency
CC	during any stage of an organism's development. The nucleic acids are also
CC	useful as probes and as a guideline to define new PCR (polymerase chain
CC	reaction) primers for the cloning of substantially homologous DNA
CC	sequences from other sources. The nucleic acids are also useful for
CC	determining the presence or quantity of beta-diox II nucleic acid and
CC	determining presence and amount of beta-diox II. The polypeptide is also
CC	useful for increasing or decreasing the amount of beta-diox II levels in
CC	a cell or tissue which can modify the level of vitamin A and other
CC	retinoids. Antibodies are useful for studying beta-diox II localisation,
CC	screening of an expression library to identify nucleic acids encoding
CC	beta-diox II or the structure of functional domains. The transgenic
CC	fungi, yeast, insect, animal or plant cells, seeds, tissues or whole-
CC	organisms have improved nutritional quality or physiological condition
CC	and accumulate important metabolites of carotene/retinoid pathways such
CC	as vitamin A aldehyde and retinoic acid, beta carotene or take up beta
CC	carotene from the medium. Expression systems encoding beta-diox II are
CC	useful in the study of beta-diox II activity. Identification of cDNAs
CC	encoding beta-diox I and II allows the physiological characterisation of
CC	mammalian carotene/retinoid metabolism. Vitamin A production in crops and
CC	microorganisms can be achieved by transforming the organisms or crops
XX	with the cDNAs.
SQ	Sequence 2134 BF; 625 A; 448 C; 441 G; 620 T; 0 other.
Query Match	8.1%; Score 250.8; DB 22; Length 2134;
Best Local Similarity	50.8%; Prod No 2 2e-51;
Matches 789; Conservative	0; Mismatches 712; Indels 51; Gaps 6

[illegible][illegible]





XX AA053074.  
 AC  
 XX  
 XX 02-JUN-1994 (first entry)  
 DE Human retinol binding protein receptor cDNA  
 XX  
 XX Retinol binding protein receptor; retinitis pigmentosa; ss.  
 OS Homo sapiens.  
 XX  
 XX W09323538-A.  
 XX  
 XX 25-NOV-1993.  
 XX  
 XX 14-MAY-1993: 93MO-0504586.  
 XX  
 XX 15-MAY-1992: 92US-0883539.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 PA  
 PI Haverl CO, Eriksson U, Simon A;  
 XX  
 XX WPI: 1993-386570/48.  
 DR  
 DR P-PSDB: AAR44618.  
 XX  
 XX  
 XX New retinol binding protein receptor and homologue coding nucleic acid molecule - useful for diagnosis and treatment of retinoid linked pathological conditions, for hybridisation in stringent conditions and treating retinitis  
 PT  
 PT  
 PS Claim 5; Page 26, 44pp, English.  
 CC  
 CC This sequence may optionally have a nucleotide extension at its 5'-region. The possible extensions are given in AA053071-73.  
 CC  
 XX  
 XX Sequence 1265 BP: 379 A, 310 C, 210 G, 363 T; 3 other;

Query Match 3.48; Score 104.6; DR 14; Length 1265;  
 Best local Similarity 47.98; Pred. No. 1.6e-15;  
 Matches 515; Conservative 0; Mismatches 521; Indels 39; Gaps 6;

UY 636 GGCACCTCTCCACACACATATGACAGTGGCTGCAATATTCGACATGGGTCTTCAT 695  
 DB 2 GGCACCTCTCCACACATATGACAGTGGCTGCAATATTCGACATGGGTCTTCAT 61  
 YY 636 TCTTGATTAAGGAGCAACAAATATTTCTTTAAGATCCCTCTCTGACACAAAA 755  
 DB 62 TGCAGAAAAATTTTCAATTGGCTACACATTTAAACATCCG-----ACCACTGCA 112  
 YY 756 AGAAAGAGAAATCTGTTTAAACACGAGTACTGTCATCCCTCTGCTC 815  
 DB 113 AGCAGATACAGCAATCTGATTAAGCAAGTACAGATGCTGCAATTCCTCCAGTGA 172  
 YY 816 CCTGCTCCACCAAGACACACACCTTTGAAATCAGCAAGAAATTAATTTGTTTCA 875  
 DB 173 CCGATTAATACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 232  
 YY 876 AGAGATGATTTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 932  
 DB 233 GATATACATCTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 292  
 YY 933 GAATGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 992  
 DB 293 CAATCATATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 352  
 YY 993 CAGAAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1052  
 DB 353 CAAAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 412  
 YY 1053 CCAATATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1112  
 DB 413 TCACATCAACACCTATGATGAATATGAGGTTTCTATTTCTATTTCTATTTCTAT 472

UY 1113 CAATAGCT---TGACAGATATGTTTAAATTAATAATAATAATAATAATAATAATA 1169  
 DB 473 ATTTCACCTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 942  
 YY 1170 CAAATAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 1229  
 DB 533 CAAAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 992  
 YY 1221 CAAATAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1280  
 DB 593 TATTCACAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 652  
 YY 1281 TGAATAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1345  
 DB 653 AATTCCTGACAGTACAGAGCTATGCTCCCTGCAAGCTTACCTTACCTTACCTTAC 712  
 YY 1336 -----ATAGACTGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1388  
 DB 713 TCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 772  
 YY 1389 CAAATAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1448  
 DB 773 TCGATATCCACTTCCCTTCAATCACTTTGTTCCAGATAGGCTGTGATGCTCAAT 842  
 YY 1449 AACAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1508  
 DB 833 AACAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 892  
 YY 1509 TCCAGAGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1568  
 DB 893 TCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 952  
 YY 1569 TG---AGCAATTAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCA 1625  
 DB 953 AGCAGCAATTAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 1012  
 YY 1626 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
 DB 1013 TGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1067

RESULT 14  
 ABL29861  
 ID ABL29861 standard; DNA: 1978 BP.  
 XX  
 XX ABL29861;  
 AC  
 AC  
 XX 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 41016.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 OS Drosophila melanogaster.  
 XX  
 XX W0200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 PD 23-MAR-2001: 2001MO-0509231.  
 PF  
 PF 23-MAR-2000: 2000US-191637P.  
 PR 11-JUL-2000: 2000US-0614150.  
 PP  
 PP (PERE) PE CORP NY.  
 PA  
 PA Venter JC, Adams M, Li PWD, Myers EW.  
 PI  
 PI WPI: 2001-754997/75.  
 DR  
 DR New isolated nucleic acid detection reagent for detecting 1000 or more

[illegible]

CC response studies in relevant models of retinoid/vitamin A deficiency

CC during any stage of an organism's development. The nucleic acids are also



OM protein - protein search, using SW model

Run on: July 15, 2003, 09:04:09 ; Search time 7.2 seconds

(without alignments)  
973.470 Million cell updates/sec

Title: US-10-053-192-1

Sequence: 1 MDTFNKKKEHPEHKAAY.....MHLLEHMFLEKUNLGAET 520

Scoring table: RUSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250520 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 100%  
Listing first 45 summaries

Database: A-Geneseq\_1010021\*

```

1: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1980.DAT *
2: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1981.DAT *
3: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1982.DAT *
4: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1983.DAT *
5: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1984.DAT *
6: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1985.DAT *
7: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1986.DAT *
8: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1987.DAT *
9: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1988.DAT *
10: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1989.DAT *
11: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1990.DAT *
12: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1991.DAT *
13: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1992.DAT *
14: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1993.DAT *
15: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1994.DAT *
16: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1995.DAT *
17: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1996.DAT *
18: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1997.DAT *
19: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1998.DAT *
20: /SID2/gcgdata/geneseq/gene-seq-emb1/AA1999.DAT *
21: /SID2/gcgdata/geneseq/gene-seq-emb1/AA2000.DAT *
22: /SID2/gcgdata/geneseq/gene-seq-emb1/AA2001.DAT *
23: /SID2/gcgdata/geneseq/gene-seq-emb1/AA2002.DAT *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	FR	ID	Description
1	28.5	100.0	526	21	AAV97311	Beta, Beta-carotene
2	27.21	96.3	547	22	AAV97314	Beta, Beta-carotene
3	21.07	74.6	547	22	AAV93108	Human protein seqn
4	17.15	60.7	516	22	AAV62810	Amino acid sequence
5	11.42	40.4	549	22	AAV04293	Beta-2, beta-carotene
6	11.16	39.5	579	22	AAV68872	Human PCAP polypeptide
7	11.12	39.4	532	22	AAV04292	Mouse-2, beta-carotene
8	11.02	39.0	549	22	AAV4454	Human protein seqn
9	11.02	39.0	556	22	AAV04294	Human-2, beta-carotene
10	10.71	37.9	533	22	AAV73955	Canine RPB5, Can

11	10.70	37.9	533	14	AAV44617	Human retinol bind
12	10.14	35.9	529	21	AAV97315	Beta, Beta-carotene
13	8.40	29.8	523	22	AAV71747	Neoxanthin cleavage
14	8.40	29.8	620	22	AAV94291	Protoporphyrin beta-ca
15	8.40	29.8	620	22	AAV62841	Active site sequence
16	5.76	20.4	357	14	AAV44618	Human retinol bind
17	5.41	19.2	570	22	AAV12066	Arabidopsis thaliana
18	5.20	18.4	596	22	AAV12065	Arabidopsis thaliana
19	3.73	13.2	205	22	AAV73956	Canine mutant RPB5
20	3.11	11.0	538	22	AAV34790	Arabidopsis thaliana
21	3.10	11.0	538	22	AAV04785	Arabidopsis thaliana
22	3.07	10.9	501	21	AAV31317	Arabidopsis thaliana
23	3.07	10.9	538	21	AAV31316	Arabidopsis thaliana
24	3.07	10.9	539	21	AAV31315	Arabidopsis thaliana
25	2.76	9.8	605	22	AAV73958	Neoxanthin cleavage
26	2.76	9.8	605	22	AAV73959	Neoxanthin cleavage
27	2.63	9.3	577	22	AAV04786	Arabidopsis thaliana
28	2.63	9.3	577	22	AAV94416	Arabidopsis thaliana
29	2.60	9.2	599	22	AAV04784	Arabidopsis thaliana
30	2.60	9.2	599	22	AAV93311	Arabidopsis thaliana
31	2.47	8.7	297	22	AAV14652	Neoxanthin cleavage
32	2.39	8.5	604	22	AAV04788	Neoxanthin cleavage
33	2.39	8.5	604	22	AAV73309	Neoxanthin cleavage
34	2.37	8.4	589	22	AAV91182	Arabidopsis thaliana
35	2.33	8.2	517	21	AAV31334	Arabidopsis thaliana
36	2.33	8.2	595	21	AAV31333	Arabidopsis thaliana
37	2.33	8.2	595	21	AAV04783	Arabidopsis thaliana
38	2.33	8.2	595	22	AAV73296	Neoxanthin cleavage
39	2.31	8.2	544	22	AAV73303	Neoxanthin cleavage
40	2.30	8.1	512	22	AAV04787	Neoxanthin cleavage
41	2.16	7.7	404	22	AAV14467	Neoxanthin cleavage
42	2.07	7.3	583	22	AAV04782	Arabidopsis thaliana
43	1.97	7.0	446	22	AAV73307	Neoxanthin cleavage
44	1.76	6.2	431	21	AAV31335	Arabidopsis thaliana
45	1.33	4.7	110	22	AAV14465	Neoxanthin cleavage

## ALIGNMENTS

RESULT 1  
AAV97311 standard: Protein; 526 AA.  
ID AAV97311:  
XX AAV97311:  
XX 07-JAN-2001 (first entry)  
XX Beta, Beta-carotene-15,15-dioxygenase.  
XX DE Beta, Beta-carotene-15,15-dioxygenase; Vitamin A beta-carotene;  
XX KM transformation; fruit, vegetable, developmental disorder;  
XX KM optunaomological disorder; antibody; detection; quantification;  
XX KM treatment; therapy.  
XX OS Gallus gallus.  
XX PN EP1031627-A1.  
XX 30-AUG-2000.  
XX 17-FEB-2000; J00000-010289.  
XX 22-FEB-1999; 99EP-0103382.  
XX (Hoff) HOFFMANN LA PROTEIN 6 ON AG F  
XX Neumann H, Proger P, Friedlein AM, Witter GM, Weggen W, Wyss A;  
XX Wyss M;  
XX WPI; 2000-551036/51  
XX N-PSDB; AAA53888.  
XX

beta-carotene-15,15'-dioxygenase protein, nucleic acids and antibodies, useful for production of vitamin A from carotene and gene therapy of ophthalmological disorders

Claim 1: Page 12 13; 37pp; English.

Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bCOD are useful as primers or probes for specific amplification and/or detection of the gene that encodes bCOD, for isolation of related sequences in other organisms, for determining bCOD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bCOD gene. The nucleotide coding bCOD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bCOD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bCOD are used for detection/quantification of bCOD in immunoassays.

Sequence 526 AA:

Query Match 100.0%; Score 2825; DB 21; Length 526,  
Best Local Similarity 100.0%; Prod. No. 2 2e 247,  
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 METTFNNKKEHEPEPKAEVQGLPTWLGVLIRNKGKMHIIIGDTRKYNHMEGLALLHSF 60
DB 1 METTFNNKKEHEPEPKAEVQGLPTWLGVLIRNKGKMHIIIGDTRKYNHMEGLALLHSF 60
QY 61 TFKNGEYVYSKYLRSTNYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
DB 61 TFKNGEYVYSKYLRSDTYNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 120
QY 121 TDNCLINIMKTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 180
DB 121 TDNCLINIMKTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 180
QY 181 HNMGTSTVYKRPRTKVTFRPSSVPEKPKKSKFHELEVVN-STPSPLDQSYHSFGI 240
DB 181 HNMGTSTVYKRPRTKVTFRPSSVPEKPKKSKFHELEVVN-STPSPLDQSYHSFGI 240
QY 241 TENYIVFIEQPEKLDIVKLATAYIRGVNNAASCLSTFKEDKTFHFVDKRTKEVSTKFT 300
DB 241 TENYIVFIEQPEKLDIVKLATAYIRGVNNAASCLSTFKEDKTFHFVDKRTKEVSTKFT 300
QY 301 DALVVIHHINAYEEDGHVFDIVAYPDNSLYDMFYKLKIDKDFEVNNKLTSLPTCKRPVY 360
DB 301 DALVVIHHINAYEEDGHVFDIVAYPDNSLYDMFYKLKIDKDFEVNNKLTSLPTCKRPVY 360
QY 421 ATEVQMSPVPTKIAKLNVQTKVELHNGEHHFWSEYIVVSPGAAEEELVVLTVVVSF 480
DB 421 ATEVQMSPVPTKIAKLNVQTKVELHNGEHHFWSEYIVVSPGAAEEELVVLTVVVSF 480
QY 481 PNRAPFLIIDAFTFKELGPATVNMHMDLMCFPONDGAFFE 526
DB 481 PNRAPFLIIDAFTFKELGPATVNMHMDLMCFPONDGAFFE 526

```

RESULT 2

AAV97314  
ID AAV97314 standard; Protein: 506 AA.

AAV97314:

01-JAN-2001 (first entry)

Beta, beta-carotene-15,15'-dioxygenase.  
Beta, beta-carotene-15,15'-dioxygenase; vitamin A; beta-carotene; transformation; fruit; vegetable; developmental disorder; ophthalmological disorder; antibody; detection; quantification; treatment; therapy.

Gallus gallus.

EP1031627-A1.

30-AUG-2000.

17-FEB-2003; 2009EP-0102289.

22-FEB-1999; 99EP-0103382.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woygon W, Wyss A;

Wys M;

WPI: 2000-551036/51.

Beta, beta-carotene-15,15'-dioxygenase (bCOD) is used for enzymatic conversion of beta-carotene to vitamin A and for raising specific antibodies. Nucleotides encoding all or part of bCOD are useful as primers or probes for specific amplification and/or detection of the gene that encodes bCOD, for isolation of related sequences in other organisms, for determining bCOD levels in humans (to identify subjects requiring vitamin A supplementation) and for detecting mutations in the bCOD gene. The nucleotide encoding bCOD may also be used to transform cells, particularly plant cells, to increase their vitamin A contents (especially in fruits and vegetables) and in gene therapy of subjects who have mutated or deleted forms of the bCOD gene and thus low vitamin A levels and susceptibility to developmental or ophthalmological disorders. Antibodies raised against bCOD are used for detection/quantification of bCOD in immunoassays.

Sequence 506 AA:

```

QY 10 EHHPEPKAEVQGLPTWLGVLIRNKGKMHIIIGDTRKYNHMEGLALLHSF 69
DB 1 EHHPEPKAEVQGLPTWLGVLIRNKGKMHIIIGDTRKYNHMEGLALLHSF 69
QY 79 PSKYITSSVYVNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 129
DB 79 PSKYITSSVYVNCNIEANRIVVSEPTMAYDPCKNIFAKAFSYLSTIPEF 129
QY 130 KTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 189
DB 130 KTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 189
QY 121 KTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 180
DB 121 KTGDYVATSETNTRIKIDPQITFLIDKVDYSKYVAVNATISPHRDSAGN 180
QY 190 DKGPRTKYVLPKIPSSVPEKPKKSKFHELEVVN-STPSPLDQSYHSFGI 240
DB 190 DKGPRTKYVLPKIPSSVPEKPKKSKFHELEVVN-STPSPLDQSYHSFGI 240
QY 250 QPEKLDIVKLATAYIRGVNNAASCLSTFKEDKTFHFVDKRTKEVSTKFT 309
DB 250 QPEKLDIVKLATAYIRGVNNAASCLSTFKEDKTFHFVDKRTKEVSTKFT 309
QY 310 NAYEEDGHVFDIVAYPDNSLYDMFYKLKIDKDFEVNNKLTSLPTCKRPVY 369
DB 310 NAYEEDGHVFDIVAYPDNSLYDMFYKLKIDKDFEVNNKLTSLPTCKRPVY 369

```



PA (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.  
 XX  
 PI Von Lintig J. Voigt K:  
 XX WPI: 2001-441713/47.  
 DR N-PSDB: AAH42170.  
 XX  
 PT Dioxigenase (beta-diox) protein that cleaves beta carotene to form  
 PT vitamin A aldehyde, and polynucleotides encoding them useful for  
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide  
 PT  
 XX  
 PS Claim 7: Page 86-87; 87pp; English.  
 XX  
 CC The present sequence represents a beta-carotene dioxigenase (beta-diox);  
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form  
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic  
 CC plants. The transgenic plants have improved nutritional quality or  
 CC physiological condition and accumulate vitamin A aldehyde and can take  
 CC up beta-carotene from the medium. Expression systems encoding beta-diox  
 CC are useful in the study of beta-diox activity. Identification of cDNAs  
 CC encoding beta-diox allows the physiological characterization of  
 CC mammalian vitamin A metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 SQ Sequence 516 AA:  
 Query Match 60.7%; Score 1715, DB 22, Length 516;  
 Best Local Similarity 59.7%; Pred. No. 1,le-146;  
 Matches 310; Conservative 87; Mismatches 114; Indels 8; Gaps 5:  
 OY 1 METFNKKKEHPPEIKAEVQGLPTWLOQVILRNGPCHMTIGDTKYNHMPDGLALHSF 60  
 DB 1 MGYDYGNKKKEHPPEITEVKSSIPREVQGITLKNQGMHVSVEITNYHMFDMALLHSF 60  
 OY 61 TFKNGEYYSKYRSTNYCNIEANRIVYSEGTATYEDPCNITFAKFSYLSHTPEF 120  
 DB 61 AINKEGVYTSRTRIGDTYSNMQANIVYSEGTATYEDPCNITFAKFSYLSHTPEF 120  
 OY 121 TTNCLINIMKTGGDYAISETNIRKIDPYLETLKRVYSKYAVANLAISHGVDSACN 180  
 DB 121 TDNCGNNIIRKYNDGFHATSETNIRKIDPYLETLKRVYSKYAVANLAISHGVDSACN 180  
 OY 181 ILMNGTSIVDKGRKRYVLEKIP--SSYPEKSKCKRHLEVSCISPSRLDPSYHSFG 239  
 DB 181 SYSNGTIAKGGKTKRWLEKVPSPSPSPSP---LKSAAVCTLPSPSLTPSYHSFG 237  
 OY 240 ITENYVIEIPEPKLDIVKLATAYIRGVNMASSCLSEKEDKTFHFVDKTKKSTKRY 299  
 DB 238 MTNVPFIFEPGLKLDILKMATAYIRFVSNASCMKRPEDSTLILHLENTKRYATKRY 297  
 OY 300 TDAIVLVIHINANEEDGHVVDIVARSDNSLYMEFLKLIKDFEVNKKLISIPCKKRV 359  
 DB 298 TDAAITVHYQNAFEDDGHVVDIVAYDNNLVEFFELNFKFTMGATN-LYKPKFTFFV 356  
 OY 360 VPILOYKDAEVSNSLVKLP-TSATAVKKGDSIYCOPTLIGGIELPRVNYNKKRYK 418  
 DB 357 FPLI--SIQGTGEDGLKLYTTASAKKEDKIMCGEVLGSGVELPRINTFNCKRTY 414  
 OY 419 VATEVWSPFTRIKALNVQREVLHMGEDHCPSEPIFVPSPAKREDECVLTGVV 478  
 DB 415 SYMCVDSESPVATRIKFPADTRQQLKEMKGDPRFASFPVIFPPGAVDEDDGVLTVIN 474  
 OY 479 SEFNKAPFLLIDAKTFKELGRATVNEHMLDHGMFTP 517  
 DB 475 NKPIQSGFLLVLEAKSFKEAACLDELHMDHGMFTP 513  
 RESULT 5  
 AAU04293  
 ID AAU04293 standard; Protein: 549 AA.  
 XX

AC AAU04293:  
 XX  
 DT 23-OCT-2001 (first entry)  
 XX  
 DE Zebra-2, beta-carotene dioxigenase (Beta-diox II).  
 XX  
 KW Beta-carotene dioxigenase; beta-diox II; zebra-2; beta-carotene;  
 KW lycopene; beta-apocarotenal; beta-ionone; apolygonal; grain seed; corn;  
 KW oil seed; palm; chick pea; diapaestic; therapeutic; ribozyme; zebrafish;  
 KW retinoid/vitamin A deficiency; beta-diox II; transgenic nutrition;  
 KW carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.  
 XX  
 OS Brachydanio rerio.  
 XX  
 IN W020048163-A2.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 27 DEC 2000, 2000MW-EF13-73.  
 XX  
 PE 24-DEC-1999, 99EP-015895.  
 XX  
 FR 20 MAR 2000, 2000EP-015892.  
 XX  
 PA (GREE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.  
 XX  
 PI Von Lintig J. Voigt K:  
 XX  
 DR WPI: 2001-425657/45.  
 XX  
 DR N-PSDB: AAS07195.  
 XX  
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxigenase  
 PT (beta-diox II) protein that cleaves beta carotene and lycopene to yield  
 PT beta-apocarotenal and beta-ionone, and apolygonals, respectively.  
 PS Claim 2: Fig 14; 116pp; English.  
 CC The sequence represents the amino acid sequence of zebra 2, beta-carotene  
 CC dioxigenase (beta-diox II). Beta-diox specifically cleaves beta-carotene  
 CC and lycopene to form beta-apocarotenal and beta-ionone, and  
 CC apolygonals, respectively. The DNA is useful for transforming grain  
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;  
 CC edible seeds or seeds with edible parts e.g. chick peas; potatoes;  
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or  
 CC ribozyme type therapeutic agents and for detecting any abnormality of  
 CC endogenous beta-diox II. The beta-diox II specific antisense  
 CC oligonucleotides derived from the DNA sequence are useful for dose  
 CC response studies in relevant models of retinoid/vitamin A deficiency  
 CC during any stage of an organism's development. The nucleic acids are also  
 CC useful as probes and as a guideline to define new PCR (polymerase chain  
 CC reaction) primers for the cloning of substantially homologous DNA  
 CC sequences from other sources. The nucleic acids are also useful for  
 CC determining the presence or quantity of beta-diox II nucleic acid and  
 CC useful for increasing or decreasing the amount of beta-diox II levels in  
 CC a cell or tissue which can modify the level of vitamin A and other  
 CC retinoids. Antibodies are useful for studying beta-diox II localization,  
 CC screening of an expression library to identify nucleic acids encoding  
 CC beta-diox II or the structure of functional domains. The transgenic  
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole  
 CC organisms have improved nutritional quality or physiological condition  
 CC and accumulate important metabolites of carotenoid/retinoid pathways such  
 CC as vitamin A aldehyde and retinoic acid. beta-carotene or take up beta  
 CC carotene from the medium. Expression systems encoding beta-diox II are  
 CC useful in the study of beta-diox II activity. Identification of cDNAs  
 CC encoding beta-diox II and II allows the physiological characterization of  
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 SQ Sequence 549 AA:  
 Query Match 40.4%; Score 1142.5, DB 22, Length 549:





AA00429 standard; Protein: 532 AA.  
AAU04292:  
23-OCT-2001 (first entry)  
Mouse-2, beta-carotene dioxygenase (beta-diox II).  
beta-carotene dioxygenase; beta-diox I; mouse-2; beta carotene,  
lycopene, beta apocarotenal; beta ionone; apolycopenal; grain seed; corn;  
oil seed; palm; chick peas; diagnostic; therapeutic; ribozyme;  
retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;  
carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop.  
Mus musculus.  
MO200148163-A2.  
05-JUL-2001.  
27-DEC-2000; 2000MO-EF13273.  
24-DEC-1999; 99EP-0125895.  
20-MAR-2000; 2000EP-0105822  
(GEE-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH  
Von Lintig J., Vogt K;  
WP1: 2001-475657/45.  
N-PSDB: AAS07194.  
Novel isolated mouse, human, zebrafish beta-carotene dioxygenase  
(beta-diox II) protein that cleaves beta carotene and lycopene to yield  
beta-apocarotenal and beta-ionone, and apolycopenals, respectively -  
Claim 2; Fig 14; 116pp; English.

The sequence represents the amino acid sequence of mouse-2, beta-carotene  
dioxygenase (beta-diox II). Beta-diox specifically cleaves beta-carotene  
and lycopene to form beta-apocarotenal and beta-ionone, and  
apolycopenals, respectively. The DNA is useful for transforming grain  
seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm;  
edible seeds or seeds with edible parts e.g. chick-peas; potatoes;  
carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
for diagnostic and/or therapeutic purposes and for preparing antisense or  
ribozyme type therapeutically agents and for detecting any abnormality of  
endogenous beta-diox II. The beta-diox II specific antisense  
oligonucleotides derived from the DNA sequence are useful for dose  
response studies in relevant models of retinoid/vitamin A deficiency  
during any stage of an organism's development. The nucleic acids are also  
useful as probes and as a guideline to define new PCR (polymerase chain  
reaction) primers for the cloning of substantially homologous DNA  
sequences from other sources. The nucleic acids are also useful for  
determining the presence or quantity of beta-diox II nucleic acid and  
determining presence and amount of beta-diox II. The polypeptide is also  
useful for increasing or decreasing the amount of beta-diox II levels in  
a cell or tissue which can modify the level of vitamin A and other  
retinoids. Antibodies are useful for studying beta-diox II localisation,  
screening of an expression library to identify nucleic acids encoding  
beta-diox II or the structure of functional domains. The transgen-  
ic fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole  
organisms have improved nutritional quality or physiological condition  
and accumulate important metabolites of carotene/retinoid pathways such  
as vitamin A aldehyde and retinoic acid. beta-carotene or beta-ca-  
rotenoide from the medium. Expression systems encoding beta-diox II are  
useful in the study of beta-diox II activity. Identification of cDNAs  
encoding beta-diox I and II allows the physiological characterisation of  
mammalian carotene/retinoid metabolism. Vitamin A production in crops and  
microorganisms can be achieved by transforming the organisms or crops  
with the cDNAs.

Query Match	39.4%	Score 1112.5	Pos 272	Length 542
Best local Similarity	43.5%	Pred. No. 5,46-92		
Matches 228	Conservative 94	Mismatches 175	Indels 27	Gaps 10
QY	10	EEHPEPIKAEVQGLPTMLDGLVLLNPGDMITLDTIKYINWIFSTALIHSTFKKAEVY	69	
Db	19	EEFLSASAPVNGHIEPEMLNLYLLVPGVEEFGKRYINWIFSTALIHSTFKKAEVY	78	
QY	70	KSKYLSADLYNNLANKVNSDGLMAYDGLKNHAKAKVLSLSTFELTHRLNIM	129	
Db	79	KSFVLDSTSTYKANSAGPVLISFSTTALPRPKSTIFFEKKVFFP--LTMTLNNVY	166	
QY	130	KTGDDVYATSETNFLPKIDPQTELEKVDYSKYVAVNAATSHRYNSAGNLLMSTIV	189	
Db	137	QYKGDVYMTSTENFNKKYIDIMLEPKTEKVDMSKELAVNAIAIDHYDPKGLAVNMINSYQ	196	
QY	190	DKRRTYVLFKIPSSVPERKKKSGFKLELVCSIPRSISLDPSYHNSGTENYVTE	249	
Db	197	PRG-SCYNIIRV---PKKKEGETIHQAVLCSIASTEKKKSYHNSGMIKNILFE	252	
QY	250	QPEKLDIVKATATVTCGVNMASTLSEFKKEKTFWHEVLEKIKENSLEYTHALVTHN	309	
Db	253	QYFKMLTWLIISSKIDPEKFLALGIMELVLEFENVALRELGLDGLMAYDGLNLYD	316	
QY	310	NATVEEPGVHPTVAVVPRNSLYCMYLYKK-----LPTVTVNNKLLSTPTVYVYD	362	
Db	313	NATVEEPGVHPTVAVVPRNSLYCMYLYKK-----LPTVTVNNKLLSTPTVYVYD	366	
QY	363	QYDKDAEVSQNLVLDL--LSALAVKAKLSLSTVYDGLDLE	413	
Db	369	DVSVLAAEKKNLSPLSTYSASAVKQYKDELMQSPENLDHDLFEEDGLFEPQINVGPN	426	
QY	414	KKKYLYVATLEVWMSVPLKIAKLVNLTKEVGLNEMFENFMSSTFVYSSTAAEDELVY	474	
Db	429	KKYSFYYGGGPR-HLVGSLIKVULTKTLRVNMEFEGFYSFVSEVYRVVADDELSDVLL	487	
QY	474	TCVYVSEPRKARFLLDLAKTFKELFSGATVNVEMHLLDMLPRP	517	
Db	488	SVVITPNQSHSNFLLVLDKRSFTHLGAEVAVYDMLRPTPHDTVPR	531	
RESULT 8				
AAAB93458				
ID	AAH93458	standard: Protein: 539 AA.		
AC	AAH93458.			
XX				
XX	24-JUN-2001	(first entry)		
CT				
XX		Human protein sequence SEQ ID NO:12718.		
DE				
XX		Human; primer: detection; diagnosis; antisense therapy; gene therapy.		
XX				
XX		Homo sapiens.		
XX				
PN	EP1074617-A2.			
XX	07-FEB-2001.			
PD				
XX	28-JUL-2000; 2000EF-0116126.			
XX				
XX	24-JUL-1999; 99JP-0246036.			
XX	27-AUG-1999; 99JP-0002531.			
XX	13-JAN-2000; 2000JP-0118776.			
XX	02-MAY-2000; 2000JP-0183767.			
XX	09-MAY-2000; 2000JP-0241894.			
XX				
PA	(HELI-) HELIX RES INST.			
XX				
XX	Oka T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;			
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;			
XX				

DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX  
 PS Claim 8: SEQ ID 12718; 2537bp + CD ROM; English.  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNA-estly without any specialised methods. AAH3156 to AAH1568 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92449 to  
 CC AAH95894 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX  
 SO Sequence 539 AA:  
 Query Match 39 0%: Score 1102.5; P: 22; Length 539;  
 Best Local Similarity 44.3%; Pred. No. 4,48-91;  
 Matches 232; Conservative 85; Mismatches 178; Indels 29; Gaps 10;  
 QY 10 EHPPIKAEVGGOLPTMLOGVLLNPGGMMHTTGGTKYNNHFDGLALHSTFKNGEYVY 69  
 DB 28 EAPPGIISARWGHCFKMLNLSLKICKCKPEKNDKMHFEDMALHCKMAKGVITY 87  
 QY 70 FSKYPTPTTYNTNANFVSEPTMAVPECEKRIKAKKS YLHIIIEITELIN 127  
 DB 88 RSKPLQSTTYKANSKKNIVISRTTALPPPKNVFPPSPRLPKAAAMTNTNVN 147  
 QY 128 IMKTDVVYATSETNFKRIDQTLLEKLVYSYAVANLAISSHPHYDSAGNLLNGTS 187  
 DB 148 VYRKGDDVYLCETETFMKMKVDELTLEKEDVMSKFTAVN-CATNHPHYDPGTAYMNGNS 207  
 QY 188 IVDPKPTVYVLFRTISSVPEKEKKSCTPKH EVAGSTPSEGLDFSYTHSGITENIVE 247  
 DB 208 PEPYCP-PSYKVLVPR--PFRVCTGGTTHVAGVCTASTKPKSPSYHSRMLNVIIF 269  
 QY 248 IEOPKRLIVKLATVYIRGVNMASTLSFHEEKTWTFIEVDEKIKKEVSIKMYIALVLYH 307  
 DB 264 IEPQKMNIMKATISKITGKAFSGTISMEVGCNIRFHVEKFTQGLLPSYVSKVYVIRH 323  
 QY 308 HINAVEEDGVVHIVAVRDSIVMPYIKKIK--DFEVNKLITSTTKRFVET 362  
 DB 324 QINAEEDGVVHIEEDQGLNFTLEVYCIQNLFRAGETLIDVHNSAKSPF--FFVATL 381  
 QY 363 QVDKAEVGSN-VKLP-TSATAVNKKKGSIVCGPELIRB-----RIELPKVAVYD-VNG 414  
 DB 382 NWSLNAPEODNLSPLSTYSASAVKQADGIWCSHEDNLHDLKEGGIEFQIYTHPFSJ 441  
 QY 414 KRYKYVAVTEVQWSPVPTKIAKLVNCTEVLHWRDHPWSPSEPTFVSPFAFEDESVL 473  
 DB 442 KRYHFFYGGGPR-HI-VGDSLIKVNIV-----WPRDGPYPRPEVFAVAPGTFNPDGVIL 494  
 QY 474 TCVVVSERKAFELLILAKTKFKELGATVAVENHLLHGMFIP 517  
 DB 495 SVVITPNQNSNELLVLDAKNFPIGRRAEVQVQMPYGHGTFIP 538

RESULT 9  
 AAU04294  
 ID AAU04294 standard; Protein; 556 AA.  
 AC  
 XX  
 AC AAU04294;  
 DI  
 XX  
 DI 23-OC1-2001 (first entry)  
 DE  
 XX  
 DE Human-2, beta-carotene dioxygenase (beta-diox II).  
 KW  
 KW beta-carotene dioxygenase; beta-diox II; human-2; beta-carotene;  
 KW lycopene; beta-apocarotenal; beta ionone; apolycopenal; grain seed; corn;  
 KW oil seed; palm; chick-pea; diagnostic; therapeutic; riboflavin;  
 KW retinoid/vitamin A deficiency; beta-diox II; transgenic; nutrition;  
 KW carotene; retinoid pathway; vitamin A aldehyde; retinoic acid; crop.  
 XX  
 US Home sapiens.  
 XX  
 XX WC2001042163-A2.  
 XX  
 XX 05-JUL-2001.  
 XX  
 XX 27-DEC-2000; 2000WB EP13273.  
 XX  
 XX 24-DEC-1999; 1999F-01-5895.  
 XX  
 XX 20-MAR-2000; 2000EP-0105822.  
 XX  
 PA (GPRF-) GPRFN-VATION PLANT/ENRICHMENT/LOGIC GMBH.  
 XX  
 XX Von Lintig J, Vogt K;  
 PI  
 XX  
 DR WPI: 2001-425657/45.  
 DE  
 XX N FSDR: AAS07196.  
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase  
 PT (beta-diox II) protein that cleaves beta carotene and lycopene to yield  
 PT beta-apocarotenal and beta-ionone, and apolycopenals, respectively -  
 XX  
 XX Claim 4, Fig 14, 11epg; English.  
 CC The sequence represents the amino acid sequence of human 2, beta-carotene  
 CC dioxygenase (beta-diox II). Beta diox specifically cleaves beta-carotene  
 CC and lycopene to form beta-apocarotenal and beta-ionone, and  
 CC apolycopenals, respectively. The DNA is useful for transforming grain  
 CC seeds, e.g. corn; oil seeds, e.g. brassica seeds; edible seeds e.g. palm;  
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes;  
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or  
 CC ribozyme type therapeutic agents and for detecting any abnormality of  
 CC endogenous beta-diox II. The beta-diox II specific antisense  
 CC oligonucleotides derived from the DNA sequence are useful for dose  
 CC response studies in relevant models of retinoid/vitamin A deficiency  
 CC during any stage of an organism's development. The nucleic acids are also  
 CC useful as probes and as a guideline to define new PCR polymerase chain  
 CC reaction primers for the cloning of substantially homologous DNA  
 CC sequences from other sources. The nucleic acids are also useful for  
 CC determining the presence or quantity of beta-diox II nucleic acid and  
 CC determining presence and amount of beta-diox II. The polypeptide is also  
 CC useful for increasing or decreasing the amount of beta-diox II levels in  
 CC a cell or tissue which can modify the level of vitamin A and other  
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,  
 CC screening of an expression library to identify nucleic acids encoding  
 CC beta-diox II or the structure of functional domains. The transgenic  
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole  
 CC organisms have improved nutritional quality or physiological condition  
 CC and accumulate important metabolites of carotene/retinoid pathways such  
 CC as vitamin A aldehyde and retinoic acid. beta-carotene or take up beta-  
 CC carotene from the medium. Expression systems encoding beta-diox II are  
 CC useful in the study of beta-diox II activity. Identification of cDNAs  
 CC encoding beta-diox I and II allow the physiological characterisation of  
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and

CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.

XX Sequence 556 AA:

Query Match 39 0% Score 1162, Pos 22, Length 556;  
 Host Local Similarity 44.1% Pred. No. 4,6e-91;  
 Matches 231; Conservative 87; Mismatches 183; Indels 23; Gaps 9;

```

UY 10 EEHEPPIKAEVQGLPTWLGVLKNGGMMHTIGIRKYNHWELGLALLSHFTKKEGVY 69
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 39 EEARGISARVWGHPFRLWGLSLRLIGPKFERGKPKYNIMFPMALHGFMAKQTVY 98
UY 70 RSKYLRSDTYMNCNENRPIVVSFEGTMAVPPCKNIFAKAFS--YLSHTIPERTQCLIN 127
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 99 RSKFLQSDTYKANKAKRRIYISSEGLTALPDCKNFERMSFELPGRMAAATDDTNN 158
UY 128 IMKTGDYATSEINFRKIDPOTLELDKVDYSKYVAANLATSHPHYDSAGNLMGTS 187
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 159 YVRKRGDYLLCTETINFMNKVDILETEKTEKVDMSKFLAVGATAHHPYDPDGTAYNMGNS 218
UY 188 IVDGPTKRYVLFKTPSSVPEKKEKSCFKHLEVVGSTPSSLLQFSYHSFGITENYIV 247
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 219 FGPRG-FSYKVIKRP---PEVLDGETIHGVQVICSISTAEKKEKPSYHSFGITRYITF 274
UY 248 IEGPFLDIKATATAYIRGVNMAASCLSFHEKEDKTMHFVDKTKKEVSTKPYTDALVYH 307
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 275 IEGPLKNNLMKRIATSKIRGKAFSDGISMEPCNTREHVEKRTGQLDPCRYYSKPEVTH 334
UY 308 HINAEHKGHVVDIVAYRONSILYDMFLKLLDK-----DFEVNKLTLTSPCKRRVVL 362
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 435 QINAEHQGVVIMNCGQDMSPRLFYVQIQNLKACGGLNQVHNSAKSP--RPFVLEPL 392
UY 463 QYDQDAEVSNLVKLP-TSATAVKEKDSIYCOPELLCE-----GIELPRVND-YNG 413
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 393 NVSLNADEGNLSPLESTASAVKQADGTTICSHENLHODELEKEGIEFPQIYYDRSG 452
UY 414 KRYRYVATEVQMSPVTKIAKLNQVTEKYLHMGECHEWSPPIVPSPDAREDEGVVL 473
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 KRYHEFGGGR-RLVDSLSLKVDVNNKTLKVRDEGFYSEPVFAPAGTNEDEGVIL 511
UY 474 TGVVSEPNKAPFLILDATKFEKIGRATVNMHLDLGMFIP 517
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 512 SVVITPNQNSNELLVDAKNEFEIGRAEYVQMPYGFHGTIP 555

```

RESULT 10

AAB73955 standard; Protein: 533 AA.

XX AAB73955:

DT 29-MAY-2001 (first entry)

XX Canine RPE65.

XX Canine dog: RPE65, canine retinal pigment epithelium 65; RPE65;

XX congenital stationary night blindness; CSNB; mutation detection.

XX Canis familiaris.

XX 056201114-H1.

XX 13-MAR-2001.

XX 30-AUG-1999; 9905-0385259.

XX 06-OCT-1998; 9805-0103219.

XX (CGR ) CORNELL RES FOUND INC.

XX Aquitro GD, Acland GM, Ray K;

DB WPI: 2901-265168/27.

DB N-PSDB: AAB739560.

XX Novel nucleic acid molecule encoding canine retinal pigment epithelium  
 PT 65, where presence of mutation in one or both alleles is indicative of  
 PI a carrier of, or dog affected with congenital stationary night  
 XX blindness

PS Claim 2: Fig 1A-1B; 25pp; English.

CC The present sequence is canine retinal pigment epithelium 65 (RPE65). The  
 CC gene encoding this protein is useful for identifying dogs which are  
 CC genetically normal, or are carriers of, or affected with congenital  
 CC stationary night blindness (CSNB). Carriers for CSNB are identified  
 CC through the detection of an AACA deletion at nucleotides 487-490 in the  
 CC canine RPE65 gene. This allows a breeder to eliminate the carrier from  
 CC the breeding stock or to breed carriers with genetically normal dogs.

XX Sequence 533 AA:

Query Match 37.9% Score 1071; Pos 22; Length 533;  
 Host Local Similarity 42.2% Pred. No. 3,1e-88;  
 Matches 222; Conservative 32; Mismatches 188; Indels 24; Gaps 12;

```

UY 2 ETIFRNKKEPPIKAEVQGLPTWLGVLKNGGMMHTIGIRKYNHWELGLALLSHFT 61
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17 EIV-----EELSSPLAHVGRLEHMLSSLEGGCHDFEVSSEDEYHLEFMAHRED 71
UY 62 FKRGEEVYRKVLRSDTYMNCNENRPIVVSFEGTMAVPPCKNIFAKAFS--YLSHTIP 121
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 72 FKRGEEVYHPPPIVVSFEGTMAVPPCKNIFAKAFS--YLSHTIPSPSYHPPV FVT 129
UY 122 DNCLINIMKTGSEYVATSEINFRKIDPOTLELDKVDYSKYVAANLATSHPHYDSAGN 181
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 130 DNALVNVYVPGVYVYATETNPTIKINPTIETKLVNDGNNVSGATAHHPYDPDGTAY 189
UY 182 LMGTSIVDGGRIRKIVLEKLPSSVPEKKEKSCFKHLEVVGSTPSSLLQFSYHSFGIT 241
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 190 YNIGNCEGKNESIAVNVKIF--FLQADKEEDISKEVVGVQPSNDREKPSYVHSPIT 246
UY 242 ENYIVTEPPEKLDIVATAY-IPVNNMAASCLSFHEKEDKTMHFVDKTKKEVSTKPY 400
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 247 PNYIVVETIVKINLKLFLSSWSIWGANIMDETHSNLMVNLHAIKAKKAYLNMYRT 406
UY 301 DALVYHINAEYEDGHVFDIVAYRP-NSLYTMPIYKILKIIFE--VNNKLISDPYCK 456
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 307 SSENLFPHHINTYETNEFLIVDLCWKGFEEVYVYATLANIPENMFVKKNAKAPQPEVR 466
UY 357 PFVVPRLQYDQDAEVSNLVKLP-TSATAVKEKDSIYCOPELLCE-----GIELPRV 410
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 367 RSVLEPLNIDK-ADTGKMLVTLPPNTVATILRSDETTWLEPEVLSGPGWATFEPQINQK 425
UY 411 YNCKRYRYVVAIEVQMSPVTKIAKLNQVTEKYLHMGECHEWSPPIVPSPDAREDEG 470
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 426 YGGRPYTAYAGGLNHF-VPDRICKLNIVKIKETIMWQGEDVSPSEPTFVSHUALDEIDG 484
UY 471 VVLTCVYVSEPNKAP-FLILDATKFEKIGRATVNMHLDLGMFIP 515
DB 11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 485 VVLSVYVSPGAGOKPALLLLNKKDLSEVARAEVETINIPVTPHGLE 530

```

RESULT 11

AAR44617 standard; Protein: 533 AA.

XX AAR44617:

XX 02 JUN 1994 (first entry)

XX Human retinol binding protein receptor.

XX Retinol binding protein receptor; retinitis pigmentosa.

OS Homo sapiens.  
 XX MO9323538-A.  
 XX 25-NOV-1993.  
 XX  
 XX 14-MAY-1993: 93WO-US04586.  
 XX  
 XX 15-MAY-1992: 92US-0883539.  
 XX  
 XX (LUDWIG) LUDWIG INST CANCER RES.  
 XX  
 XX Bavi CO, Eriksson U, Simon A;  
 XX WPI: 1993-386570/48  
 XX N-PSDB: AA053070.  
 XX  
 XX New retinol binding protein receptor and homologue coding nucleic  
 XX acid molecule - useful for diagnosis and treatment of retinoid  
 XX linked pathological conditions, for hybridisation in stringent  
 XX conditions and treating retinitis  
 XX  
 XX Claim 15, Page 21-24; 44pp; English.  
 XX  
 XX The protein (SFO, ID NO. 1) may be used to diagnose and treat  
 XX retinoid linked conditions such as inability to accumulate  
 XX retinoids in the eye, leading to retinitis pigmentosa.  
 XX  
 XX Sequence 533 AA;

Query Match 37.9%; Score 1070; DB 14; Length 533;  
 Best Local Similarity 42.0%; Pred No. 3.9e-88;  
 Matches 221; Conservative 92; Mismatches 189; Indels 24; Gaps 12;

2 ETITFNPKFHEPEIKAEVQGLFTW:GCVLLFNGFDMHTIGCTKRYNHMEDLALNSFI 61  
 17 EIV-----EELSPILAHVTGRIPMLTGLSLKRCQPGI FEVASEPFIHIFGQALHKKFD 71  
 62 FKKEGYYSKSKLRSTYKNCLEANRIYSEGGTMAPDPCCKNIFAKAFSLSHITPEFT 121  
 72 FKGGHTVIRPFIITAVFAFMEKPIVITETGTAFEDPKNIFSPFVSFVS- EVT 129  
 122 DNLINIMKTGDYVATSENFIRKIDVLTLETLKVCYSKVAVNAATISHIHVSANI 181  
 130 DNALVNIYVGEYVACTETNITVNPETLETIQVGLCNVSVNGATAPHIENQGV 189  
 182 LMGSTIVDKRTKYVLRKISSVEKEKKSKTRHLEVRSTPSSLTQPSYHSFST 241  
 190 YNIGNCFGNFSLAANIYKIP--PLQADKEDPISKSEIYVOFPCSDRFKPSYHSFST 246  
 242 ENYIVFIEDPFIKDIYKLTAY-IRGVNMASSLSFHKEDKTMFIHFVDRKTRKESYKFT 300  
 247 PNIVIVETPVIKINLFKFLSSMSLMGANYMDFESNETMGVLAHIAKKRRKYINNKRT 306  
 301 DALVYHHINAYEEDGHVVDIAVARD-NSLYDMFYKLLDKDF--VNNKLISITLTK 356  
 307 SEFNLFHHINTEYEHFELVDLQCKKSEFVYNYLYLANLRFNMHFFVKNMFKAPQEV 366  
 357 HHVPLVYVTKFAEVSNLVKLP-TSATAVKHFVNSLYVFFELVPS-----LPIEVNLT- 410  
 367 RYVLPINLTK-ADTKCNITVLPNTATATLSTHTLMEFVYIFSGPQAFEPFQINQK 425  
 411 YNGKTKVYVATEVQSPVETRIAKLVGCTKEVIMGSETHWPSSEFIVPSFCLAFEEES 470  
 426 YGKRKYTVAYIGLNIHF-VPDRICKINVTKTRETWVQDEPDSPPITVSHPTALEEDG 484  
 471 VITVVSSEPNKAP-FITLIDAKTFEIGRATVVENHIGIHSNF 515  
 485 VHSVVSVPAGQCKPAYITLNAKDLSEVAPAFVEINIPVTFHDLF 530

RESULT 12  
 AA97315

ID AAY97315 standard; Protein; 529 AA.  
 XX  
 XX AAY97315;  
 AC  
 XX 03-JAN-2001 (first entry)  
 DT  
 XX  
 XX Beta, Beta-carotene-15,15-dioxygenase.  
 XX  
 XX Beta, beta carotene 15,15 dioxygenase; vitamin A; beta-carotene;  
 XX transformation; fruit; vegetable; developmental disorder;  
 XX ophthalmological disorder; antibody; detection; quantification;  
 XX treatment; therapy.  
 XX  
 XX Bos taurus.  
 XX  
 XX EP1031627-A1.  
 XX  
 XX 30-AUG-2000.  
 XX  
 XX 17 FEB-2000; 2000EP 0102289.  
 XX  
 XX 22-FEB-1999; 99EP-0103382.  
 XX  
 XX (HOPF ) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 XX Bachmann H, Brugger R, Friedlein AM, Wirtz GM, Woggon W, Wyss A;  
 XX WISS M;  
 XX WPI: 2000-551036/51.  
 XX  
 XX Beta, beta-carotene 15,15'-dioxygenase protein, nucleic acids and  
 XX antibodies, useful for production of vitamin A from carotene and gene  
 XX therapy of ophthalmological disorders  
 XX  
 XX Disclosure, Page 19-20, 37pp; English.  
 XX  
 XX Beta, beta-carotene-15,15'-dioxygenase (bcoo) is used for enzymatic  
 XX conversion of beta-carotene to vitamin A and for raising specific  
 XX antibodies. Nucleotides encoding all or part of bcoo are useful as  
 XX primers or probes for specific amplification and/or detection of  
 XX the gene that encodes bcoo, for isolation of related sequences in  
 XX other organisms, for determining bcoo levels in humans (to identify  
 XX subjects requiring vitamin A supplementation) and for detecting  
 XX mutations in the bcoo gene. The nucleotide encoding bcoo may also  
 XX be used to transform cells, particularly plant cells, to increase  
 XX their vitamin A contents (especially in fruits and vegetables) and  
 XX in gene therapy of subjects who have mutated or deleted forms of the  
 XX bcoo gene and thus low vitamin A levels and susceptibility to  
 XX developmental or ophthalmological disorders. Antibodies raised  
 XX against bcoo are used for detection/quantification of bcoo in  
 XX immunoassays.  
 XX  
 XX Sequence 529 AA;

Query Match 35.9%; Score 1014.5; DB 21; Length 529;  
 Best Local Similarity 42.3%; Pred No. 4.1e-83;  
 Matches 225; Conservative 91; Mismatches 184; Indels 53; Gaps 20;

10 EHHPEIKAEVQGLFTW:GCVLLFNGFDMHTIGCTKRYNHMEDLALNSFI 61  
 17 EIV-----EELSPILAHVTGRIPMLTGLSLKRCQPGI FEVASEPFIHIFGQALHKKFD 71  
 62 FKKEGYYSKSKLRSTYKNCLEANRIYSEGGTMAPDPCCKNIFAKAFSLSHITPEFT 121  
 72 FKGGHTVIRPFIITAVFAFMEKPIVITETGTAFEDPKNIFSPFVSFVS- EVT 129  
 122 DNLINIMKTGDYVATSENFIRKIDVLTLETLKVCYSKVAVNAATISHIHVSANI 181  
 130 DNALVNIYVGEYVACTETNITVNPETLETIQVGLCNVSVNGATAPHIENQGV 189  
 182 LMGSTIVDKRTKYVLRKISSVEKEKKSKTRHLEVRSTPSSLTQPSYHSFST 241  
 190 YNIGNCFGNFSLAANIYKIP--PLQADKEDPISKSEIYVOFPCSDRFKPSYHSFST 246  
 242 ENYIVFIEDPFIKDIYKLTAY-IRGVNMASSLSFHKEDKTMFIHFVDRKTRKESYKFT 300  
 247 PNIVIVETPVIKINLFKFLSSMSLMGANYMDFESNETMGVLAHIAKKRRKYINNKRT 306  
 301 DALVYHHINAYEEDGHVVDIAVARD-NSLYDMFYKLLDKDF--VNNKLISITLTK 356  
 307 SEFNLFHHINTEYEHFELVDLQCKKSEFVYNYLYLANLRFNMHFFVKNMFKAPQEV 366  
 357 HHVPLVYVTKFAEVSNLVKLP-TSATAVKHFVNSLYVFFELVPS-----LPIEVNLT- 410  
 367 RYVLPINLTK-ADTKCNITVLPNTATATLSTHTLMEFVYIFSGPQAFEPFQINQK 425  
 411 YNGKTKVYVATEVQSPVETRIAKLVGCTKEVIMGSETHWPSSEFIVPSFCLAFEEES 470  
 426 YGKRKYTVAYIGLNIHF-VPDRICKINVTKTRETWVQDEPDSPPITVSHPTALEEDG 484  
 471 VITVVSSEPNKAP-FITLIDAKTFEIGRATVVENHIGIHSNF 515  
 485 VHSVVSVPAGQCKPAYITLNAKDLSEVAPAFVEINIPVTFHDLF 530

QY 244 YIVFIDQPKLIDVLAATAY INGVNMASTL SFHKEDK TWRFHYDRTKKEVSTKFT 300  
 DB 248 YIVFETPVKINLKEKLSWSLWGANMDC FESFTETMTQVWLHIDKRRKKLNNKYPT 297  
 QY 301 DALVLYHHINAYEEDGVNFDIVAYRD-NSLYDMF--YLKKLDDKE---VNNKLSIPT 354  
 DB 298 SFPMFLPHHINTYEENGLVIDLCWKGFEVYVFTLYLANLENNVEYKKAKRKAPQPE 357  
 QY 355 CKRRVVFLOYDKDAEVGSNLVKLP-TSATAV--KEKDGSTYQPELCEG---IELPV 407  
 DB 358 VRRVVLPLINDK-ADTKNLVTLPNTATATILCSDEFTTITLPEVVLFGSPROAPEFPI 416  
 QY 408 NYD-YNGKKYVVAIFVQMSPVPTKIAKLVQIKVELH--WGEDHGWSEPIFVSPDA 464  
 DB 417 NYQYCGCKPYTAAAGLDLNIH-VPRDLCKLVNKTKEWETVMOEPDSTSEPTLVSHPA 475  
 QY 465 REEDGVVLCVAVSEPNKAP-FLILIDAKTFKELGRA--TVNVEMLDLHGK 515  
 DB 476 LEEDGVVLSVAVSPGAGCKPAVILLINAKULSEVAKAEFTVEINIPVIFHGLE 529  
 RESULT 14  
 ID ABB71747 standard; Protein: 620 AA.  
 AC ABB71747:  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polyprotein SFQ ID NO 42033  
 KM Drosophila developmental biology cell signalling; insecticide;  
 US pharmaceutical.  
 OS Drosophila melanogaster.  
 PN W0200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-0509231.  
 PE 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI: 2001-656860/75.  
 DR N-PSDB; ABL15850.  
 XX  
 PE New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosures: SFQ ID NO 42033, 21pp + Sequence Listing, English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/publ/indbnd\\_jpat\\_sequences](http://wipo.int/pub/publ/indbnd_jpat_sequences)  
 XX Sequence 620 AA;  
 SQ

Query Match 24 88; Score 849.5; DP 22; Length 620;  
 Best Local Similarity 34.68; Pident. No. 4126-67;  
 Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;  
 QY 9 KEHPPEIKAEVQGLPTLWGLVLLPNDGMDTIDPTKINIMHDLALHSPFNKQVY 68  
 TB 50 FRFIVPTEDHSHSHIPKWIQNSLDNPGQSNWVDMTPGHLPPSSALDHRAKINQVY 109  
 QY 69 YKSTYLSISQYN-NIENKLVSEFGLMAYFDEKNIKAKASYSLSHTPEE TQNLIN 127  
 TB 110 YQNPFDVTEFLPKRPSQRIVVTEPFAAPDCHSIFDR---FAAFPPSSGINSMS 166  
 QY 128 IMKTGDYVATSETNFRKIDPOTLETLKVDKSKVAANLATSHPHYASNLIMGTS 187  
 DB 167 TYPFGDDYVTFTEPFRHPIINPCTLATEARLCTDVGAVNNTSHRHVLPSTVNLGIT 226  
 QY 188 LVDKGRTKYLKLPSSVPEKKKSCFKHLEVCSLPKSLQSYNSPGLDENVYVF 247  
 CB 227 MTPSGPAYTLL-----SFPHGEV---MFEIAHVAVTLPEFMKIPQYMTPEGLDHPYV 278  
 QY 248 IEEPFKLDIYELATAYIPGVNMASTLSEHREKTWRFHYDRTKKEVSTKFTTALVLYH 407  
 DB 279 VEDPLVSLTEYTKAOLGQNLNLSLTKWETDMLPHLLIDKVSQKLVQI YSEKAPVTH 467  
 QY 308 HINAVFTGQVNVVAVVPRVETQMTVYPTVQAFVYVKKVYVYVYVYV 462  
 DB 338 IINCERDGHVVDICSYRNFEMINMYLFAI-ANNQININVAITLFRCHDLREVLGL 496  
 QY 363  
 TB 337 PRASIANPGLVKSSTAGLSANVVSFLPNSVSAVNIYMNINRDLATAHLSRPRDOR 456  
 QY 370 VG-----NLVKKLP-----SALAVEKEDGSIYVPELEDFLELPPVNVVYVYKRYVYV 419  
 DB 457 KGRYEENLVNLVIMESQAEAFQGTNG-IGLRRPEMLCWMGCEPTRIYFHRMKKNYVF 515  
 QY 420 YATFVQMSPV-PTKIAKLVNQTKEVHLHMGDHCWSEPIFVSPDAREHDSVVLTCVV 478  
 DB 516 YAISSDVAVNPGLTLLKVDVWNSCLTWCEENVYSEPIFVSPDPKSDIDGVILLASVYL 575  
 QY 479 SEPN-KAPFLILIDAKTFKELGRATVNVEMHLD-----LHGMEFIP 517  
 DB 576 GGLNDKRYVGLIVLIDAKTMTLHGK-----DPRHINQVPRKGLDHPFAP 617  
 RESULT 14  
 ID AAU04291  
 AC AAU04291 standard; Protein: 620 AA.  
 DT 23-OCT-2001 (first entry)  
 XX Drosophila beta-carotene dioxygenase (beta-diox 1).  
 KM Beta-carotene dioxygenase; beta-diox 1; Drosophila; beta-carotene;  
 KM lycopene; beta-apocarotenal; beta-ionone; apolytrypenol; grain seed; corn;  
 KM oil seed; palm; chick pea; diagnostic; therapeutic; lycopene;  
 KM retinoid/vitamin A deficiency; beta-diox 1; transgenic; nutrition;  
 KM carotene/retinoid pathway; vitamin A aldehyde; retinoic acid; crop;  
 XX Drosophila melanogaster.  
 PN W0200148163-A2.  
 XX 05-JUL-2001.  
 PD 27-DEC-2000; 2000WO-EP13273.  
 PE 24-DEC-1999; 99EP-0125495.  
 PR 20-MAR-2000; 2000EP-0105822.  
 XX (GREE-) GREENOVATION PLANTZENBIOTEC/NOVIGIE GMBH.  
 PA

XX Von Lintig J, Vogt K;  
 XX MPI: 2001 425657/45.  
 DR N-PSDB: AAS07180.  
 XX  
 PT Novel isolated mouse, human, zebrafish beta-carotene dioxygenase  
 PT (beta-diox II) protein that cleaves beta-carotene and lycopene to yield  
 PT beta-apocrotene and beta-ionone, and apolycopene, respectively  
 XX  
 PS Disclosure: Fig 6; 116pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of *Drosophila* beta-  
 CC carotene dioxygenase (beta-diox II). Beta-diox II specifically cleaves beta-  
 CC carotene and lycopene to form beta-apocrotene and beta-ionone, and  
 CC apolycopene, respectively. The DNA is useful for transforming grain  
 CC seeds, e.g. corn; oil seeds, e.g. Brassica seeds; edible seeds e.g. palm,  
 CC edible seeds or seeds with edible parts e.g. chick-peas; potatoes;  
 CC carrots, sweet potatoes, sugar beets, banana. The polypeptide is useful  
 CC for diagnostic and/or therapeutic purposes and for preparing antisense or  
 CC ribozyme type therapeutic agents and for detecting any abnormality of  
 CC endogenous beta-diox II. The beta-diox II specific antisense  
 CC oligonucleotides derived from the DNA sequence are useful for dose  
 CC response studies in relevant models of retinoid/vitamin A deficiency  
 CC during any stage of an organism's development. The nucleic acids are also  
 CC useful as probes and as a guideline to define new PCR (polymerase chain  
 CC reaction) primers for the cloning of substantially homologous DNA  
 CC sequences from other sources. The nucleic acids are also useful for  
 CC determining the presence or quantity of beta-diox II nucleic acid and  
 CC useful for increasing or decreasing the amount of beta-diox II levels in  
 CC a cell or tissue which can modify the level of vitamin A and other  
 CC retinoids. Antibodies are useful for studying beta-diox II localisation,  
 CC screening of an expression library to identify nucleic acids encoding  
 CC beta-diox II or the structure of functional domains. The transgene  
 CC fungi, yeast, insect, animal or plant cells, seeds, tissues, or whole  
 CC organisms have improved nutritional quality or physiological condition  
 CC and accumulate important metabolites of carotene/retinoid pathways such  
 CC as vitamin A aldehyde and retinoic acid, beta-carotene or take up beta-  
 CC carotene from the medium. Expression systems involving beta-diox II are  
 CC useful in the study of beta-diox II activity. Identification of cDNAs  
 CC encoding beta-diox I and II allows the physiological characterisation of  
 CC mammalian carotene/retinoid metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 XX Sequence 620 AA;  
 SO  
 Query Match 29.8%; Score 840.5; DB 22; Length 620;  
 Best Local Similarity 34.5%; Prof No 3.2e-67;  
 Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;  
 OY 9 KEHPRKAEVQGGPLTGLGVLRLNPGMHTTDDTYNNIMFGLALLHFTKNGEY 68  
 DB 50 EREIVDPITGHHSGHPKMGISILRNKSGSKWVLMFPHGLPESALLHFAALNKGVT 109  
 OY YSKYLRSIDYNCNLEARNIVSEFGTMAVPPDCKNIFAKAFSYLSRTIPF-DNDCIN 127  
 DB 110 YONRFVDETILKKNSAGRIIVTRGTAAVAPDPCISIFDR---FAITFRPISGPDNMSIS 166  
 OY 128 IMKGGDYATSETFNFIKIDQLEILDKVSKYVAVNLATSHENYASGNTINMST 187  
 DB 167 IPEFGQVYTFPTEPFMHPTNCTATFARICTTTPVGVVNHSHPHVLPSTVYNNJGT 226  
 OY 188 IYDKGRTKYVLFKISSVDEKSKKSKFHLVGVSIIPSPSLQPSYHSGTENTIVE 247  
 DB 227 MRPSPAVTIL---SPPHTEQ---MEDIHVAATFPCFPLHNGHTFGLTHFVFI 278  
 OY 248 IEQPRKLDIVKATAYIRGVNMSASLSEHKEDKTMFHFVDKRTKEVSTKYATALVYH 307  
 DB 279 VEQPLSVSLTEYIKAGLGQGNLSACTLKMFEDEBPLFHLIDVNSKLVQI-VESAPFYIH 347  
 OY 308 HINATFEGHVVVDIVATPDNSLITPMFLTKIKLIDKPFVNNKLSTIPCK--PRVAVI--- 362

DB 338 IINCEERGGHVVVDIVATPDNSLITPMFLTKIKLIDKPFVNNKLSTIPCK--PRVAVI--- 366  
 OY 363 -----OY-----DKDAE 369  
 DB 397 PRASIAKRLGVKSFSLAGISAPQVSRITMKHSVSQADITTPYRNGKATANGESPKKDAK 456  
 OY 370 VG-----SNVKLPT-----SATAVKEKLSIGCEPELICE-VIELPKNVY--YNNKATAYV 419  
 DB 457 RGRYEENLVNLTVEGSAQAEAFQGTNG-IQLFREMIDMGCETPRIYERMYKKNRYKP 515  
 OY 420 YATVEQWSPV-PTTAKTINVTCTFVLIHMGCPHMPSPETIYVSPDAPPPPVAVI-IVVV 478  
 DB 516 YAISSVVAVNDGTTILKVTVMKSGTITWCFEYVSPPTIYVSPKSPDPAVVIASVVL 575  
 OY 479 SRPN-KAPLLITDAKTRKELGRATVYEMHLD-----LHGMPIP 517  
 DB 576 GGINPQYVGLVIVDAKMTLELSFG----PFTNNGPVPCIHGWAP 617  
 RESULT 15  
 ID AAG62841 standard; Protein; 620 AA.  
 XX  
 AC AAG62841;  
 XX  
 DT 17-SEP-2001 (first entry)  
 XX  
 DE Amino acid sequence of a beta-carotene dioxygenase (beta-diox).  
 XX  
 KW Beta-carotene dioxygenase; beta-diox; beta-carotene; vitamin A aldehyde;  
 KM transgenic plant.  
 XX  
 OS *Drosophila* melanogaster.  
 XX  
 IN W020014162-A2.  
 PD 05-JUL-2001.  
 XX  
 IT 22-DEC-2000; 2000MO-BP13144.  
 XX  
 PP 24-SEP-1999; 99EP-0125895.  
 PP 20-MAR-2000; 2000EP-010-822.  
 XX  
 PA (GREF-) GREENOVATION PFLANZENBIOTECHNOLOGIE GMBH.  
 XX  
 PI Von Lintig J, Vogt K;  
 XX  
 DR MPI, 2001-441713/47.  
 DR N-PSDB: AAH42156.  
 XX  
 PT Dioxygenase (beta-diox) protein that cleaves beta-carotene to form  
 PT vitamin A aldehyde, and polynucleotides encoding them useful for  
 PT producing transgenic bacteria, fungi, plants expressing the polypeptide  
 XX  
 PS Claim 7; Fig 6; 87pp; English.  
 XX  
 CC The present sequence represents a beta-carotene dioxygenase (beta-diox)  
 CC polypeptide. Beta-diox specifically cleaves beta-carotene to form  
 CC vitamin A aldehyde. Beta-diox is useful for producing transgenic  
 CC plants. The transgenic plants have improved nutritional quality of  
 CC physiological condition and accumulate vitamin A aldehyde and can take  
 CC up beta-carotene from the medium. Expression systems encoding beta-diox  
 CC are useful in the study of beta-diox activity. Identification of cDNAs  
 CC encoding beta-diox allows the physiological characterisation of  
 CC mammalian vitamin A metabolism. Vitamin A production in crops and  
 CC microorganisms can be achieved by transforming the organisms or crops  
 CC with the cDNAs.  
 XX  
 XX Sequence 620 AA;  
 SO  
 Query Match 29.8%; Score 840.5; DB 22; Length 620;

Best Local Similarity 34.68; Pred. No. 3,2e-67;  
Matches 203; Conservative 99; Mismatches 189; Indels 95; Gaps 18;

```

QY      9 KEHEPEPIKAEGVQGLPTWLOGVLLRNGPGMHTIGDTKYNHMFQDGLALHSTFKNGEYV 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      50 EREIVDPLEGHSHIHKWICGSLRNGPGMWKYGDMTFCGLFDGSLALHRAIRNGRYT 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      69 YRSKYLRSDTYNCNIFANNRIYVSEFGTMAFPDCKNIFAKAFSYLSHTIPEF-TDNCLIN 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      110 YQNRKFDTEITRKRRSRQRIYVTEFGTAAPDPCHSIEFR---FAAIFPDGSDTDSMTS 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      128 IMKTGDDYATSETNFIKKIDPQLETLIDKVDYSKYVAANLATSHPHYDSAGNILMGTS 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      167 IYPPGDDGYTFTEIFPMHRIINPCTLATFAPICITDEGVYVNHSHPHVLPSGTIVYNLGTT 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      188 IVDKGRITKYVLFKIPSSVPEKEKKSCFKHLEVVCSIPRSLLQPSYHSFGITENYIVF 247
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      227 MTRSRPAYTII-----SPPHGQ---MFEIAHVAVALLPCRWKILPGYMHTFGLTDHYFVI 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      248 IEOPEKLDIVKIAVATIGVNMASCLSFHKEDKTWFHVDKRTKKKEVSTKFTYDALVLYH 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      279 VEOPIVSLSLTFEYTKAQIGGQNLSSQCLKMFEDPPTLPHLIDRVSGKLVOI-YESEAFYLIH 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      408 HINAVEEGHAYVPIIVAVPONSLYDMFYLLKLLKDEEVNKKLISIPCK--RFVPL--- 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      338 IINCFFPDGHVVDICSYPNDEMIMTYLFAI-ANMOTNPVATLFEQGPLREVLPLCTI 396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      363 -----QY-----DKDAE 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      397 PPASTAKGLVKSSTLACLSAPQVSKTMAHVSQYADLIIMPIINCKQATAGEESPRKDAK 456
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      370 VG----SNLYKLIPT----SATAVEKDGSIYQPEILCE-GIELPVNYD-YNGKKRYV 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      457 KGRYEENLVMVLVTMGSSQAEAFQGIING-IOLKPEMLQMGCEPRIYERYGKNRYRF 515
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      420 YATEVQMSPV-PTKIATLNVQTEKVLHMGEDHQMSPPIFVPSPDAREDEGVVLTGVV 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      516 YAISSDVDAVNPCTLIIVDVWNKSCLTWCENVPSPPIFVPSDPKSEDDGVILASMYL 575
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      479 SEPN-KAPFLLILDAKTFKELGRATVVENMID-----LHGMFIP 517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      576 GGLNDRYVGLIIVCAKIMTEIGRC---DPIHTNGPVPRKCIHGMFAP 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 15, 2003, 09:27:36  
Job time : 75 secs





```

Db      307 SPENLFIHINTYEDHEFLVDLCWKGFEVYVNSYLANLRKEMWEKKAKARAPQPEVK 366
      357 KFYVPLDYLUKAEVGSNMLALP-1SALAWEKDSYICQPELCEG----TELPKYNVD- 410
      367 KYVPLNLIDK-ADTGKMLVTLPLNTATATILCSDETLMLEPEVLSGKQAFEPPOINYOK 425
      411 YNGKKRYVATEVQWSPVPTKLAKLWOTKEVLMHGEDHCWSPSEPIFVPSPPAREPDR 470
      426 YGCKPYTAYGLGILNH-VDPRLCKLWKTKEVWQFQPSYSEPIFVPSHPALBEDDG 484
      471 VLTGVVSPENKAP-FILLDAKFEIGRATVNEMLDLHGM 515
      485 VLVSVVSPGAGQKPAVLLIIMAKDLSEVARAEVINEIVTFGLF 530

```

## RESULT 2

P88115

Protein F53C3.12 [Imported] - *Caenorhabditis elegans*C.Species: *Caenorhabditis elegans*

C.Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

C.Accession: F88115

Anonymous, The *C. elegans* Sequencing Consortium

Science 282, 2012-2018, 1998

A.Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biolog

A.Reference number: A5000; MIM: 906613; PMID: 9819196

A.Note: see websites genome.wustl.edu/gsc/C.-elegans/ and www.sanger.ac.uk/Projects/C.-el

A.Note: Published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

C.Accession: F88115

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-556 &lt;STO&gt;

A.Cross-references: CB, Ch11, F10N A-67462.1, F10 3478489, GSPR-090062, GSPR-090062

C.Genetics:

A.Gene: F53C3.12

A.Map position: 2

## Query Match

Best local similarity 31.76; Score 778.5; DB 2; Length 556;

Matches 163; Conservative 110; Mismatches 223; Indels 19; Gaps 11;

```

      14 EPIKAEVQGLPTVLQGVLLRNGCMHTIGDTKYNHMFGLALLHSFTKNGEYVRSKY 73
      44 EFKLSTGSGVPSVLTKMLKNGMFEIDTKYQHFDGMSFLQKHEHEDGKMYSAKI 103
      74 LRSTYMCNIEANRIVSEFGTMAYPDPCKNIFAKAFSVLSHTPIPTDNLINIKTGD 133
      104 LSENKYKKNMAGQRIYSGFGTASFDPCKSIFSRFSSVQS-EGIHNNANVAFAPVD 162
      134 DYATSETNFIKIDPQTEFLDKVDYSKYVAANLASHPHYDSAGNIIMGTSTVDKGR 193
      163 GLVACTETPNKHPHPLASIDTTEVPVPSKYVALHTCTAGLPENDGVNYSRGGDAA 222
      194 TKYVLEKIPSSVPEKPKKSCFRLLEVCSIPRSLLQPSYHSFGITENYIVLEDPK 253
      223 HVTIVTKNPKML-QSDSDRS-WEHTKIGELRCESEFYPTMHSFMSNVLIMPESPIR 280
      254 LDIVKATATVIRGVNMASSCLSFHKEDKTFHFVDRKTKKEVSTKFTDALVLYHINAYE 313
      281 IDIKKELMKRPITTFEDCKMKHADQKIFILNKKTGEOVPLKLMAPEFTFHHANTE 340
      314 EDGHNVEDIAYVRDNLSDYDMFYLLKLDKDFEVNKKLTSIPTCKRFVPLQYQKDAEVSN 373
      341 KKGGLVVDYCRIBAGNFDAALLIENM-KTGNFQNDALFLPYLTKVILPLSLIDGAPGQD 399
      374 LVR---LPTSATAVKKKDSIVQPELICE-GIEEPVND-YNKKRYVYVATEV---Q 425
      400 LKLELWAKGSAIFQDQKIFLKEKRVDTISMEPFYHWEKINMKPYVYVSSVLADQ 459
      426 WSPVPTKIATINOTKRVLMGRDH---CWSPSEPIFVSPAREDEGVLTGVVSP 481
      460 KSEPLPVTKADLNGDHKVPPEPNDKQIT--GEPIFVNPESVPELDDGILLIVPWITSD 517

```

```

      482 NKAPELLILAKIFKELSPALV-NVEMHLLHGMF 515
      518 GQRPVILLKKNLTETARTTIPARIPLOFHAFY 552

```

## RESULT 3

T10688

hypocherical protein T16118.20 - *Arabidopsis thaliana*C.Species: *Arabidopsis thaliana* (mouse ear cress)

C.Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999

C.Accession: T10688

R.Bevan, M. Lennard, N. Quail, M. Harris, B. Rajandream, M.A. Barrell, R.G. Ben

submitted to the Protein Sequence Database, June 1999

A.Reference number: T10688

A.Accession: T10688

A.Molecule type: DNA

A.Residues: 1-616 &lt;BEV&gt;

A.Cross-references: EMBL: A1049915, GSPR-090062, GSPR-090062

A.Experimental source: cultivar Columbia; BAC clone T16118

C.Genetics:

A.Gene: A1SP:T16118.20

A.Map position: 4

## Query Match

Best local similarity 15.58; Score 437; DB 2; Length 616;

Matches 143; Conservative 68; Mismatches 189; Indels 104; Gaps 21;

```

      20 WQSDLEPLWMLQGVLLRNGCMHTIGDTKYNHMFGLALLHSFTKNGEYVRSKY 79
      97 VQKPTPMWMLNFTLKNRPSGLMNTGDHDEKHLFQYSFLVYLQFAGRIIAAURLLESDAY 156
      80 MCNLEANNIVSEFGTMAYPDPCKNIFAK-AEYLSHTTPE-----TANLTLINIKTGD 134
      157 KAARKHNPICVRRSS-----ETPKSVIINKPESGIGFIVPLFSGNSTLNNANVGIKIDP 212
      134 D-YTATSEINFIRK----IDPQTEFLDKVDYSKYVAANLASHPHYDSAGNIIMGTSTVD 187
      213 GRVCLTEF---QKSTLVCHETLETIGKFTYDDVLEDMIGSADITVET---EMWTL 265
      188 IVDGPRPKYVLFKIPSSVPEKPKKSCFRLLEVCSIPRSLLQPSYHSFGITENYIV 246
      256 IPELVKPEYKVVPEMASSNKP-----EVQSVPEKSSISMSGQWHSIAVLENNYV 315
      247 FIEQPKLIDIVKATAY---IRGVNMASSCLSFHKEDKTFHFVDRKTKKEVSTKFTDAL 303
      316 IPEMPLYSYKNNLLKAEPIPLYKFEWC-----PQLQJAIHWSKLI-GEVVAASVAVAY 368
      304 VLYHINAYEEDGH-----VVFEDVAYR-DNSLSDMEYLLKLDKDFEVNKKLTSIPTG 355
      369 VIFHFINAYEEDKQDQKAIIVITADCEENADRIILDMELDTL-----RSSHGRVLIQDA 424
      356 K--RFVVPLOYDKAEVGSNMLVRLPTSATAVKKKDSICQPELICEGIEEPVNDYNG 413
      425 FLSGFRFLD-----GSKYSGLEIAVEN-LEKH3-----FAMUMLSINLVIYG 465
      414 KRYVYVATEVQW-----SPVPTKIATINOTKRVLMGRDH---CWSPSEPIFVSPAREDE 446
      466 QKPYVVAQQAQPLQCNFNALSKVYIYGLIICHQYSVLNPLQFNQMPVPIVFRKRVNW 525
      447 GEHCHWSPSEPIFVSPAREDEEG 470
      526 HEHGMIPSEPIFVSPAREDEEG 549

```

## RESULT 4

S76169

hypocherical protein - *Synechocystis* sp. (strain PCC 6803)C.Species: *Synechocystis* sp.

A.Variety: PCC 6803

C.Date: 26-Apr-1997 #sequence\_revision 26-Apr-1997 #text\_change 26-Apr-1997

C.Accession: S76169

R.Kaneko, T. Sato, S. Kocani, H. Tanaka, A. Asanuma, F. Nakamura, Y. Miyajima,